

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]ch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted]on: Fri May 28 09:13:09 1999; MasPar time 43.77 Seconds
916.462 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-9
Description: (1-735) from US09191647.pep
Perfect Score: 5438
Sequence: 1 SNKNLTSFPSRIPFDTTELY.....TVHIIRQCQCEPTKSVLSEK 735

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 49.268; Variance 89.604; scale 0.550

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4486	82.5	601	5	Q20204	F40E10.4 PROTEIN (FRAG	0.00e+00
2	1651	30.4	1531	11	O88279	MEGF4.	0.00e+00
3	1551	28.5	1523	11	O88280	MEGF5.	0.00e+00
4	1424	26.2	739	4	O75094	MEGF5 (FRAGMENT).	9.79e-287
5	1188	21.8	530	5	Q24526	SLIT LOCUS ENCODING A	1.98e-232
6	578	10.6	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	6.53e-95
7	570	10.5	2447	13	O13149	NOTCH 2 (FRAGMENT).	3.70e-93
8	565	10.4	529	5	Q25058	FIBROPELLIN IA (FRAGME	4.60e-92
9	559	10.3	406	5	Q25059	FIBROPELLIN III (FRAGM	9.44e-91
10	540	9.9	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	1.32e-86
11	540	9.9	2470	11	O35516	CELL SURFACE PROTEIN.	1.32e-86
12	533	9.8	728	13	Q90656	TRANSMEMBRANE PROTEIN	4.42e-85
13	528	9.7	721	13	Q91902	X-DELTA-1.	5.41e-84
14	524	9.6	717	13	P87357	DELTAD TRANSMEMBRANE P	4.01e-83
15	521	9.6	832	5	Q99108	NEUROGENIC LOCUS DELTA	1.80e-82
16	522	9.6	1476	13	Q90285	PUTATIVE EXTRACELLULAR	1.09e-82
17	516	9.5	752	13	O42374	NOTCH RECEPTOR PROTEIN	2.18e-81
18	519	9.5	2531	5	O16004	NOTCH HOMOLOG.	4.88e-82
19	499	9.2	642	13	P79941	NOTCH LIGAND X-DELTA-2	1.05e-77
20	500	9.2	802	13	O57462	DELTAA.	6.38e-78
21	495	9.1	2352	5	O61240	HRNOTCH PROTEIN.	7.67e-77
22	492	9.0	615	13	O57409	DELTAB.	3.41e-76
23	489	9.0	1218	4	Q15816	TRANSMEMBRANE PROTEIN	1.51e-75
24	489	9.0	1218	4	O14902	TRANSMEMBRANE PROTEIN	1.51e-75
25	487	9.0	1218	4	O15122	JAGGED1.	4.09e-75
26	489	9.0	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	1.51e-75
27	477	8.8	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	5.83e-73
28	480	8.8	1372	5	P91526	SIMILARITY TO MULTIPLE	1.32e-73
29	471	8.7	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	1.14e-71
30	474	8.7	1219	11	Q63722	JAGGED PROTEIN.	2.57e-72
31	461	8.5	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	1.59e-69
32	463	8.5	1964	11	O35442	NOTCH4.	5.92e-70
33	444	8.2	1202	11	P97607	JAGGED2 (FRAGMENT).	6.87e-66
34	445	8.2	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	4.20e-66
35	447	8.2	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.57e-66
36	426	7.8	434	11	O55139	JAGGED2 PROTEIN (FRAGM	4.66e-62
37	422	7.8	1999	4	Q99940	NOTCH4.	3.29e-61
38	422	7.8	2003	4	O00306	NOTCH4.	3.29e-61
39	417	7.7	955	4	Q99466	NOTCH4 (FRAGMENT).	3.77e-60
40	408	7.5	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	3.02e-58
41	404	7.4	383	11	O70534	ZOG.	2.11e-57
42	404	7.4	383	11	Q62779	PREADIPOCYTE FACTOR 1.	2.11e-57
43	384	7.1	387	11	Q06007	NOTCH PROTEIN HOMOLOG	3.37e-53
44	378	7.0	156	5	Q26661	EPIDERMAL GROWTH FACTO	6.06e-52
45	380	7.0	473	5	Q25464	ADHESIVE PLAQUE MATRIX	2.31e-52

ALIGNMENTS

RESULT 1
ID Q20204 PRELIMINARY; PRT; 601 AA.
AC Q20204;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F40E10.4 PROTEIN (FRAGMENT).
GN F40E10.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA SMYE R.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z69792; E1346469; -.
DR PROSITE; PS01187; EGF_CA; 1.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 601 AA; 66669 MW; F8A72773 CRC32;

Query Match 82.5%; Score 4486; DB 5; Length 601;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAIC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 135 IKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAIC 194

```
Db     61 ETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    195 ETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYC 254

Db    121 EKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    255 EKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYC 314

Db    181 TKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    315 TKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDC 374

Db    241 LCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCD 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    375 LCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCD 434

Db    301 RQMSVGFKNPGAYLALDPLASDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    435 RQMSVGFKNPGAYLALDPLASDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVY 494

D     361 YIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVENSGKSDQLITKGKE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    495 YIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVENSGKSDQLITKGKE 554

Db    421 MLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCSATV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    555 MLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCSATV 614

Db    481 NFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDEKRIKCDKQKFRRHHIENECRSV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    615 NFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDEKRIKCDKQKFRRHHIENECRSV 674

Db    541 DRIKIAECNGYCGGEQNCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQCEPTKSVLSE 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    675 DRIKIAECNGYCGGEQNCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQCEPTKSVLSE 734

Db    601 K 601
          |
Qy    735 K 735
```

RESULT 2
ID 088279 PRELIMINARY; PRT; 1531 AA.
AC 088279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 30.4%; Score 1651; DB 11; Length 1531;
Best Local Similarity 39.4%; Pred. No. 0.00e+00;
Matches 259; Conservative 162; Mismatches 199; Indels 37; Gaps 25;

```
Db    748 SNKHLQALPKGIPKNVTELYLDGNQFTLVPGQ-LSTFKY-LQLVDLSNNKISSLSNSSFT 805
          |||:| ::|  || : ||||||:| :  :|:: |: : | |  :|||:|:: || |::|:
Qy      1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRL-YSLTKLDLSHNRLISLENNTFS 59

Db    806 NMSQLTTLILSYNALQCIPPLAFQGLRSLRLLSLHGNDVSTLQEGIFADVTSLSHLAIGA 865
          |:::|:|||:||| |:|: |||| || :||:||||||:| | :: |:::||::|:|:|:
Qy     60 NLTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGS 119

Db    866 NPLYCDCHLRWLSSWVKTGYKEPGIARCAGPPEMEGKLLLTTPAKKFECQGPPSLAVQAK 925
          |:||||:: |:| |:|: : |:|||||  |  :  ||||: :  | |::  :  : :|
Qy    120 NSLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATK 179

Db    926 CDPCLSSPCQNQGTCHNDPLEVYRCTCPSGYKGRNCEVSLDSCSSNPCGNGGTCHAQEGE 985
          || ||:||| | : |   :  | | | :|: | :||  :|:| ::|| | :||   ::
Qy    180 CDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQA- 238

Db    986 DAGFTCSCPSGFEGLTCGMNTDDCVKHDCVNGG-----V--C---V-D--G--IGNYTCQ 1030
            | | |  ||||  | :| ||||:  | |||     |  |   : :  :  | :| |:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db   1031 CPLQYTGRACEQLVDFCSPDLNPCQHEAQCVGTPEGPRCECVPGYTGDNCSKNQDDCKDH 1090
          ||::| |: ||: :::|:  ||||:::: |:    :  | | ||:||:||  | ||||:
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db   1091 QCQNGAQCVDEINSYACLCAEGYSGQLCEIPPAPRNSCEGTE-CQ-N--G-ANCVD-QGS 1144
           :||||: ||| | || |||  ||:|| |||||      : |: || :  | ::||  | |
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db   1145 RP-VCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANITLQVSTAEDNGILLYN 1203
             | |  ||:|| |:: :|| | :  :|| :  | :   : ||: : |:   |||||
Qy    418 SDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPLAS-D-GTITMTLRTTSKIGILLYY 475

Db   1204 GDNDHIAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVTFDQMVNLSIDGGS 1263
           ||:  :: |||:|:|:: |  |::|:| :||:  :|||  | : : | :::  | ||  :
Qy    476 GDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNP 535

Db   1264 PMTMDNFGKHYTLNSEA-P-LYVGGMPVDVNSAAFRLWQILNGTSFHGCIRNLYINN--- 1318
             ::| ||   | : :   ||:||:|:: :  | |  :: |: |: ||| :: ||:
Qy    536 VQIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPI 595

Db   1319 ELQD-FTKTQMKPGVVPGCEPCRKLYCLHGICQPNA-TP-GPVCHCEAGWGGLHCDQ 1372
           :||: : :    :  :  : |  : :| | | :|:|::  :| |||:
Qy    596 NLQQALENVNTEQSCSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDE 652
```

RESULT 3
ID 088280 PRELIMINARY; PRT; 1523 AA.
AC 088280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 28.5%; Score 1551; DB 11; Length 1523;
Best Local Similarity 36.7%; Pred. No. 0.00e+00;
Matches 241; Conservative 173; Mismatches 209; Indels 34; Gaps 17;

```
Db    739 SNRGLHTLPKGMPKDVTELYLEGNHLTAVP-KELSTFRQLTLIDLSNNSISMLTNHTFSN 797
          ||: | ::|  :| | |||||::| :  :|  :|: :  || :|||:| :  | |:|||
Qy      1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRLYSLTKLDLSHNRLISLENNTFSN 60

Db    798 MSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTLSLSHLALGIN 857
          :::||||||:|||:|||:   ||||| :|||:|:|||||| :|:::|::|||::|:|:| |
Qy     61 LTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGSN 120

Db    858 PLHCDCSLRWLSEWIKAGYKEPGIARCSSPESMADRLLLTTPTHRFQCKGPVDINIVAKC 917
          :| |||:: |:| |||: : |:|||||  |::::::|||||:   :| | : |  :: :||
Qy    121 SLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKC 180

Db    918 NACLSSPCKNNGTCSQDPVEQYRCTCPYSYKGKDCTVPINTCVQNPCQHGGTCHLSESHR 977
          : ||:||||||: |   :   | | |  :: |  |   |::|  :|| : :|| :::: |
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    978 DGFSCSCPLGFEGQRCEINPDDCEDNDCENSATCVD------------G--INNYACVC 1022
            |:| |  ||||: || | ||| :: |||:: |||            :  ||:| | |
Qy    241 --FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDC 298

Db   1023 PPNYTGELCDEVIDYCVPEMNLCQHEAKCISLDKGFRCECVPGYSGKLCETDNDDCVAHK 1082
          | :| |  |:: ::||   :| |::::|||::: :: | | ||::|: ||| : |||
Qy    299 PMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVE 358

Db   1083 CRHGAQCVDAVNGYTCICPQGFSGLFCEHPPPMVL-LQ-TSPCDQYECQNGAQCIVVQQE 1140
          |::|: |||:: :| |:|  |::| :|| || | :  | | :|:|  |  | :|:  |
Qy    359 CQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQG-ECVASQNS 417

Db   1141 P--TCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYK 1198
          :  ||:|  ||:|| |:: ::| |    :|: | :  : ::: |:: : |    |||||
Qy    418 SDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDP--LASDGTITMTLRTTSKIGILLYY 475

Db   1199 GDNDPLALELYQGHVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVMLNQTLNLVVDKGA 1258
          ||:  :: |||:|:|:|||   : |:: :||  ||||  | : :  ::  | :|| :
Qy    476 GDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNP 535

Db   1259 PKSLGKLQK-QPAVGINSP-LYLGGIPTSTGLSALRQGADRPLGGFHGCIHEVRINNELQ 1316
             : :  | :  :     ||:||:|  :  | |:  :  :: ||| : ||:
Qy    536 VQIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPI 595

Db   1317 DFK-ALPP-QS-LGVSPGCKSCT-V-CRHGLC--RSVEKDSVVCECHPGWTGPLCDQ 1366
          ::  ||   :  : |:  : |: : | :| |  ::  : :|:| :  :|  ||:
Qy    596 NLQQALENVNTEQSCSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDE 652
```

```
LT    4
ID   O75094      PRELIMINARY;      PRT;   739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;
```

```
  Query Match          26.2%;  Score 1424;  DB 4;  Length 739;
  Best Local Similarity  37.0%;  Pred. No. 9.79e-287;
  Matches   226;  Conservative  152;  Mismatches 199;  Indels  34;  Gaps  17;

Db      6 LTNYTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSL 65
          | | ||||:::|||||:|||:|||:   ||||| :||:|:|||||| :|:::|::|||:
Qy     53 LENNTFSNLTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSI 112

Db     66 SHLALGTNPLHCDCSLRWLSEWVKAGYKEPGIARCSSPEPMADRLLLTTPTHRFQCKGPV 125
          :|:|:|:|:| |||:: |:| |:|: : |:|||||  |: ::::||||:   :| | : |
Qy    113 THIAVGSNSLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKV 172

Db    126 DINIVAKCNACLSSPCKNNGTCTQDPVELYRCACPYSYKGKDCTVPINTCIQNPCQHGGT 185
            :: :||: ||:|||||: |   :   | | |  :: |  |   |::|  :|| : :|
Qy    173 PTKLATKCDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNAT 232

Db    186 CHLSDSHKDGFSCSCPLGFEGQRCEINPDDCEDNDCENNATCVD------------G-- 230
          | ::: :  |:| |  ||||: || | ||| :: ||| : |||             :
Qy    233 CKVAQAGR--FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQ 290

Db    231 INNYVCICPPNYTGELCDEVIDHCVPELNLCQHEAKCIPLDKGFSCECVPGYSGKLCETD 290
          ||:| | || :| |  |:: :: |   || |::::||||:: ::|| | ||::|: |||:
Qy    291 INSYRCDCPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETN 350

Db    291 NDDCVAHKCRHGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVL-LQ-TSPCDQYECQNGA 348
            |||    |::|: ||| | :| | |  |::| :|| || | :  | | :|:|  |  |
Qy    351 IDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQG- 409

Db    349 QCIVVQQEP--TCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDK 406
          :|:  | :  ||:|  ||:|| |:: ::| |    :|: | :  : ::: |:: : |
Qy    410 ECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDP--LASDGTITMTLRTTS 467

Db    407 DNGILLYKGDNDPLALELYQGHVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVTLNQTL 466
          ||||| ||:  :: |||:|:|:|||   : |:: :||  ||||  | : :  ::
Qy    468 KIGILLYYGDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKC 527

Db    467 NLVVDKGTPKSLGKLQK-QPAVGINSP-LYLGGIPTSTGLSALRQGTDRPLGGFHGCIHE 524
          | :||    : :  | :  :     ||:||:|  :  | |:  :  :: |||
Qy    528 FLQIDKNPVQIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISS 587

Db    525 VRINNELQDFK-ALPP-QS-LGVSPGCKSCT-V-CKHGLC--RSVEKDSVVCECRPGWTG 577
          : ||:  :: ||   : : |:  : |: : | :| |  ::  : :|:| :  :|  ||:
Qy    588 ITINEVPINLQQALENVNTEQSCSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSG 647

Db    578 PLCDQEARDPC 588
            || | |  |
Qy    648 EHCD-EKRIKC 657
```

```
RESULT   5
ID   Q24526      PRELIMINARY;      PRT;   530 AA.
AC   Q24526;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SLIT LOCUS ENCODING A PROTEIN ASSOCIATED WITH NEURAL DEVELOPMENT WITH
DE   52D EGF HOMOLOGOUS DOMAINS (FRAGMENT).
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 89077533.
RA   ROTHBERG J.M., HARTLEY D.A., WALTHER Z., ARTAVANIS-TSAKONAS S.;
RT   "slit: an EGF-homologous locus of D. melanogaster involved in the
RT   development of the embryonic central nervous system.";
RL   CELL 55:1047-1059(1988).
DR   EMBL; M23543; G514357; -.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
```

```
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
KW   NEUROGENESIS; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     530    530
SQ   SEQUENCE   530 AA;  59457 MW;  10E5764D CRC32;

  Query Match           21.8%;  Score 1188;  DB 5;  Length 530;
  Best Local Similarity  36.9%;  Pred. No. 1.98e-232;
  Matches   171;  Conservative  111;  Mismatches 159;  Indels  22;  Gaps  16;

Db      1 MKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHG 60
          : : |:|:: :  | | ::|   : :||: |:: || | | | : : | | | | ||: |
Qy    153 VSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAICETTSSRKYTCNCTPGFYG 212

Db     61 KHCEFMIDACYGNPCRNNATCTVLEEGRFSCQCAPGYTGARCETNIDDCL-G--EI--KC 115
           |||  ||||||:|| ||||| | : |||:| |  |: |  || |||||: :  |   ||
Qy    213 VHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKC 272

Db    116 QNNAT-C---I-D--G--VESYKCECQPGFSGEFCDTKIQFCSPEFNPCANGAKCMDHFT 166
          :   |   : :  :  ::||:|:|   : |  |: |:::|:  :||| | :||:
Qy    273 VDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPING 332

Db    167 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 226
          ||| | :|| | ||  ||||| |  |||||:|||||  |:| |   |:| |||   |:
Qy    333 SYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMD 392

Db    227 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 286
          | |  | :||   | :| |  ::  :||: |:|| |::|  |:   |::| :  ::: |:
Qy    393 MEYQKTDACQQSACGQGECV-ASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALD 451

Db    287 PLRTRPEANVTIVF-SSGQNGILMYDGQDAHLAVELFNGRIRVSYDVGNHPVSTMYSFEM 345
          || : ::: :|: : :::  |||:| |:|  :: ||::||::: | :|| | | |||  :
Qy    452 PL-A-SDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVK 509

Db    346 VADGKYHAVEL-LAIKKNFTLRVDRGLARSIINEGSNDYLKLT-TPM-FLGGLPVDPAQQ 402
          | ||  | : :  : :| | |::|:   : : | | :| |      | ::||||:: :|:
Qy    510 VNDGLPHRISIRTSERKCF-LQIDKNPVQIVENSGKSDQLITKGKEMLYIGGLPIEKSQD 568

Db    403 AYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQKITPGCA 445
          | :  :::|  |:|||:  : ||   :::  |    :   :|:
Qy    569 AKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCS 611


RESULT    6
ID   Q25253      PRELIMINARY;      PRT;  2653 AA.
AC   Q25253;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH HOMOLOG SCALLOPED WINGS (SCL).
GN   SCL.
OS   LUCILIA CUPRINA (GREENBOTTLE FLY) (AUSTRALIAN SHEEP BLOWFLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; OESTROIDEA; CALLIPHORIDAE;
OC   LUCILIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SS SEEKING;
RX   MEDLINE; 96400928.
RA   DAVIES A.G., GAME A.Y., CHEN Z., WILLIAMS T.J., GOODALL S., YEN J.L.,
RA   MCKENZIE J.A., BATTERHAM P.;
RT   "Scalloped wings is the Lucilia cuprina Notch homologue and a
RT   candidate for the modifier of fitness and asymmetry of diazinon
RT   resistance.";
RL   GENETICS 143:1321-1337(1996).
RN   [2]
RP   SEQUENCE OF 39-265 FROM N.A.
RC   STRAIN=SS SEEKING;
RA   CHEN Z., NEWSOME T., MCKENZIE J.A., BATTERHAM P.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE OF 39-265 FROM N.A.
RC   STRAIN=SS SEEKING;
RA   CHEN Z., MCKENZIE J.A., BATTERHAM P.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U58977; G1389670; -.
DR   EMBL; AF032672; G2654074; -.
DR   EMBL; AF032670; G2654074; JOINED.
DR   EMBL; AF032671; G2654074; JOINED.
DR   EMBL; AF032673; G2654075; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2653 AA;  285928 MW;  8F35FD2D CRC32;

  Query Match           10.6%;  Score 578;  DB 5;  Length 2653;
  Best Local Similarity  39.7%;  Pred. No. 6.53e-95;
  Matches   100;  Conservative   50;  Mismatches  75;  Indels  27;  Gaps  15;

Db    662 CHSNPCNNGATC-IDGINKYTCQCVPGFTGVHCEININECASNPCANNGVCMDLVNG-YK 719
          | ::||:| | |   :  |||| | ||| |||||  |: | ::|| ||: |:    | ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    720 CECPRGFYDPRCLSDVDECASNPCINGGRC-E------DGI-N-E-F-I----CHCPPGY 764
          | | :||    |  ::|:| :: | |||:| :      : : | : | |    | ||  |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    765 GGKRCENDIDECSS--NPCQHGGFCVDELNAFKCQCMPGYTGLKCETNIDDCINNPCANG 822
           ||:||: :: |:   |||:: | |:    :: | | ||:|| :|||||||| |  | ||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    823 GTCIDKVNGYKCVCKVPYTGQDCE--SKLD-PCA-TNRCRNDA--K--CTPSPNFLDFSC 874
          |:|:| : :| |:|:  |:|| ||  :::|     |: |:  |     | :| |  ||:|
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    875 TCKLGYTGRYCD 886
            | |::|  ||
Qy    423 KCHEGFSGPSCD 434


RESULT    7
ID   O13149      PRELIMINARY;      PRT;  2447 AA.
AC   O13149;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH 2 (FRAGMENT).
OS   FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   TETRAODONTIFORMES; TETRAODONTOIDEI; TETRAODONTIDAE; FUGU.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NAKAMURA T., TROWSDALE J.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB004829; D1021371; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE   2447 AA;  262542 MW;  3CDA4F7A CRC32;

  Query Match           10.5%;  Score 570;  DB 13;  Length 2447;
  Best Local Similarity  39.1%;  Pred. No. 3.70e-93;
  Matches   101;  Conservative   49;  Mismatches  79;  Indels  29;  Gaps  18;

Db    713 DECLPNPCQNGGSC-LDRHNGFTCVCQAGYRGVNCEKNIDECTSGPCLNQGIC-IDGLNS 770
          | || :|| | : |       :|| | :|: ||:||: || | ::|||| : | :
```

```
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    771 YTCQCVPPFAGEHCEVELDPCSSRPCQRGGVC--L-P--SAD---Y-TY----FTCRCPA 817
          : | |   | |: || ::| | :  |: || |  |    | :   : ::    : | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    818 GWQGLHCSEDVNEC-KK-NPCRNGGHCINSPGSYICKCPSGYSGHNCQTDIDDCSPNPCL 875
            :| || : :: | || ||| | | ||   ||| |:|::|::|:||:|:|||    |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    876 NGGSCVDDVGSFSCECRPGFEGEHCEIEA--D-ECA-SQPCRNGAIC-R-DYV---NS-- 924
          ||||||| : |: | ||||: |: ||| :  | |   :::|: :| | : : |   ||
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSA-CGQGECVASQNSSD 419

Db    925 FVCECRLGFDGILCDHNI 942
          | | |: || |  ||: :
Qy    420 FTCKCHEGFSGPSCDRQM 437
```

RESULT 8
ID Q25058 PRELIMINARY; PRT; 529 AA.
AC Q25058;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FIBROPELLIN IA (FRAGMENT).
OS HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BISGROVE B.W.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; L33861; G499686; -.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01186; EGF_2; 10.
DR PROSITE; PS01187; EGF_CA; 7.
DR PFAM; PF00008; EGF; 10.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 529 AA; 55543 MW; 6385F322 CRC32;

Query Match 10.4%; Score 565; DB 5; Length 529;
Best Local Similarity 39.0%; Pred. No. 4.60e-92;
Matches 99; Conservative 47; Mismatches 79; Indels 29; Gaps 17;

```
Db     99 DECASSPCQNGALC-VDQVNGYVCFCLPGFSGVHCETDIDECASSPCLNGGQC-INRINS 156
          | | :||| | |:|       | | | ||| ||||| :|| | :||||| : | : :
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    157 YECVCAAGFNGVNCQTNIDECASDPCENGG-----I--C---I-A--G--VNGYTCNCAS 201
          ::| |  ||:|  |: |||:| :  |||||     :  |   :    :  :|:| |:|:
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    202 GYTGTNCETEIDECAS-M-PCLNGGQCIEMVNGYTCQCAAGFTGVLCETDIDECASDPCQ 259
          | | :||  :: |:  : || | | || :  :|:| |::|||  |||:||:| :  ||
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    260 NGGVCTDTVNGYICSCVQGFTGSDCETN--IN-ECA-SGPCQ-N--G-GTCVDGVNG--F 309
          |||  | | : :| | |  |::|  ||    :: |   : :|| :  | | || : |:  |
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDF 420

Db    310 VCQCPPNYTGTYCE 323
          | |   ::|  |:
Qy    421 TCKCHEGFSGPSCD 434
```

RESULT 9
ID Q25059 PRELIMINARY; PRT; 406 AA.
AC Q25059;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FIBROPELLIN III (FRAGMENT).
OS HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BISGROVE B.W.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; L33862; G499688; -.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 5.
DR PFAM; PF00008; EGF; 7.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 406 AA; 43475 MW; 45E6EE2C CRC32;

Query Match 10.3%; Score 559; DB 5; Length 406;
Best Local Similarity 38.3%; Pred. No. 9.44e-91;
Matches 98; Conservative 54; Mismatches 75; Indels 29; Gaps 17;

```
Db     16 CNPNPCQNGAAC-IDQVNDYECICPPGFTGDNCETDIDVCASAPCRNGGAC-VDGVNGYT 73
          |  :|| | | |       | | | ||| | :|| :|| | ::|| | ::| |     :
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db     74 CNCIPGFDGDNCENNINECASNPCQNGGVCID-----G--V-N--GF-V----CTCQPGY 118
          | |  ||:|| ||:||::| :: |:||| |:|     :  : |  :| :    | |  |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    119 TGTLCETDIDECAS--NPCQNGGVCTDLVNMYTCDCLAGFTGSNCETNINECASNPCLNG 176
          |  ||  :: |:  |||:| | |  :   |:| | :||||:||||||::| :  | ||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    177 GACVDGVNGYVCQCLPNYTGTHCEIS--LDV-CQSM-PCQ-N--G-ATC-TN-VGGDYSC 226
          |:|||: :| | | | |:|  |||:  :|:  |   :|| :  | : | ::  ::|::|
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    227 ECPPGYTGINCEIDIN 242
          |  |::| :|: :::
Qy    423 KCHEGFSGPSCDRQMS 438
```

RESULT 10
ID Q06008 PRELIMINARY; PRT; 1203 AA.
AC Q06008;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NOTCH PROTEIN HOMOLOG 2 (MOTCH B PROTEIN) (FRAGMENT).
GN NOTCH2 OR MOTCH B.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX MEDLINE; 93178563.
RA LARDELLI M., LENDAHL U.;
RT "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL EXP. CELL RES. 204:364-372(1993).
DR EMBL; X68279; G287990; -.
DR MGD; MGI:97364; NOTCH2.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00066; notch; 1.
KW DIFFERENTIATION; NEUROGENESIS; REPEAT.
FT NON_TER 1 1
FT NON_TER 1203 1203
SQ SEQUENCE 1203 AA; 128982 MW; A5A95551 CRC32;

Query Match 9.9%; Score 540; DB 11; Length 1203;
Best Local Similarity 38.8%; Pred. No. 1.32e-86;
Matches 80; Conservative 50; Mismatches 57; Indels 19; Gaps 6;

```
Db    558 DECISKPCMNNGVCHNTQGS-YVCECPPGFSGMDCEEDINDCLANPCQNGGSC-VDHVNT 615
          | |:: ||:||::|  | :  | |:| ||| |: ||::|: | ::|| | ::| | :
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    616 FSCQCHPGFIGDKCQTDMNECLSEPCKNGGTC-------SD-------Y-VNSYTCTCPA 660
          |:| |: || || |: ::::|::  | ||| |       |:       : :||| | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    661 GFHGVHCENNIDECTE--SSCFNGGTCVDGINSFSCLCPVGFTGPFCLHDINECSSNPCL 718
          : | |||:::: ||   ::| | | |:    |:||:|: |||| |  :|::| :  |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    719 NAGTCVDGLGTYRCICPLGYTGKNCQ 744
          |:|:|||: :| |:|  ||:|  |:
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT  11
ID  O35516      PRELIMINARY;      PRT;  2470 AA.
AC  O35516;
DT  01-JAN-1998 (TREMBLREL. 05, CREATED)
DT  01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  CELL SURFACE PROTEIN.
GN  NOTCH2.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC  SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57B/6; TISSUE=THYMUS;
RX  MEDLINE; 93178563.
RA  LARDELLI M., LENDAHL U.;
RT  "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT  wide variety of tissues.";
RL  EXP. CELL RES. 204:364-372(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57B/6; TISSUE=THYMUS;
RA  HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL  SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR  EMBL; D32210; D1022953; -.
DR  PROSITE; PS00010; ASX_HYDROXYL; 22.
DR  PROSITE; PS01186; EGF_2; 27.
DR  PROSITE; PS01187; EGF_CA; 22.
DR  PFAM; PF00008; EGF; 34.
DR  PFAM; PF00023; ank; 6.
DR  PFAM; PF00066; notch; 2.
KW  GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ  SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match           9.9%;  Score 540;  DB 11;  Length 2470;
  Best Local Similarity  38.8%;  Pred. No. 1.32e-86;
  Matches   80;  Conservative   50;  Mismatches  57;  Indels  19;  Gaps   6;

Db    873 DECISKPCMNNGVCHNTQGS-YVCECPPGFSGMDCEEDINDCLANPCQNGGSC-VDHVNT 930
          | |:: ||:||::|  | :  | |:| ||| |: ||::|: | ::|| | ::| | :
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    931 FSCQCHPGFIGDKCQTDMNECLSEPCKNGGTC-------SD-------Y-VNSYTCTCPA 975
          |:| |: || || |: ::::|::  | ||| |       |:       : :||| | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    976 GFHGVHCENNIDECTE--SSCFNGGTCVDGINSFSCLCPVGFTGPFCLHDINECSSNPCL 1033
          : | |||:::: ||   ::| | | |:    |:||:|: |||| |  :|::| :  |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db   1034 NAGTCVDGLGTYRCICPLGYTGKNCQ 1059
          |:|:|||: :| |:|  ||:|  |:
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT  12
ID  Q90656      PRELIMINARY;      PRT;  728 AA.
AC  Q90656;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  TRANSMEMBRANE PROTEIN C-DELTA-1.
OS  GALLUS GALLUS (CHICKEN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC  NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=SPINAL CORD;
RX  MEDLINE; 95319507.
RA  HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT  "Expression of a Delta homologue in prospective neurons in the
RT  chick.";
RL  NATURE 375:787-790(1995).
DR  EMBL; U26590; G882412; -.
DR  PROSITE; PS01186; EGF_2; 8.
DR  PROSITE; PS01187; EGF_CA; 2.
DR  PFAM; PF00008; EGF; 6.
KW  GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ  SEQUENCE   728 AA;  79861 MW;  7439F575 CRC32;

  Query Match           9.8%;  Score 533;  DB 13;  Length 728;
  Best Local Similarity  38.7%;  Pred. No. 4.42e-85;
  Matches   79;  Conservative   40;  Mismatches  68;  Indels  17;  Gaps  12;

Db    302 HKPCKNGATCTNTGQGSYTCSCRPGYTGSSCEIEINECDANPCKNGGSCTDLENS-YSCT 360
          : |||| | |  |:   |||:| ||: |  || :|: | ::|| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    361 CPPGFYGKNCE--LSA-M-T-CADGP-CFNGGR-CTD---N--P-D-GGYSCRCPLGYSG 406
          |  || |  || :  : : | :|  | : | |::   |  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    407 FNCEKKIDYCSS--SPCANGAQCVDLGNSYICQCQAGFTGRHCDDNVDDCASFPCVNGGT 464
          :|| |::||:  :|| | : |: :  || | | :|||| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    465 CQDGVNDYSCTCPPGYNGKNCSTP 488
          | ||:  | | | ||| |  |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT  13
ID  Q91902      PRELIMINARY;      PRT;  721 AA.
AC  Q91902;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  X-DELTA-1.
OS  XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC  MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 95319507.
RA  HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT  "Expression of a Delta homologue in prospective neurons in the
RT  chick.";
RL  NATURE 375:787-790(1995).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 95319503.
RA  CHITNIS A.B., HENRIQUE D., LEWIS J., ISH-HOROWICZ D., KINTNER C.R.;
RT  "Primary neurogenesis in Xenopus embryos regulated by a homologue of
RT  the Drosophila neurogenic gene Delta.";
RL  NATURE 375:761-766(1995).
DR  EMBL; L42229; G807696; -.
DR  PROSITE; PS01186; EGF_2; 8.
DR  PROSITE; PS01187; EGF_CA; 2.
```

```
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   721 AA;  79922 MW;  028040EF CRC32;

  Query Match           9.7%;  Score 528;  DB 13;  Length 721;
  Best Local Similarity  39.2%;  Pred. No. 5.41e-84;
  Matches   80;  Conservative   39;  Mismatches  68;  Indels  17;  Gaps  12;

Db    297 HKPCENGATCTNTGQGSYTCSCRPGYTGSNCEIEVNECDANPCKNGGSCSDLENS-YTCS 355
          : || | | |  |:   |||:| ||: | :|| ::: | ::|| | ::|   : : : |
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    356 CPPGFYGKNCE--LSA-M-T-CADGP-CFNGGR-CAD---N--P-D-GGYICFCPVGYSG 401
          |  || |  ||  :   : : | :|  | :  | |::   |  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    402 FNCEKKIDYCSS--NPCANGARCEDLGNSYICQCQEGFSGRNCDDNLDDCTSFPCQNGGT 459
          :|| |::||:   ||| | ::|  :  || | |  ||:| ||: |:||| :  |||||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    460 CQDGINDYSCTCPPGYIGKNCSMP 483
          | |||  | | | ||| |  | :|
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT   14
ID   P87357      PRELIMINARY;      PRT;   717 AA.
AC   P87357;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAD TRANSMEMBRANE PROTEIN PRECURSOR.
GN   DELTAD.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97346722.
RA   DORNSEIFER P., TAKKE C., CAMPOS-ORTEGA J.A.;
RT   "Overexpression of a zebrafish homologue of the Drosophila neurogenic
RT   gene Delta perturbs differentiation of primary neurons and somite
RT   development.";
RL   MECH. DEV. 63:159-171(1997).
DR   EMBL; Y11760; E307461; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   SIGNAL; TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    717       DELTAD TRANSMEMBRANE PROTEIN.
SQ   SEQUENCE   717 AA;  79061 MW;  5CC32ECA CRC32;

  Query Match           9.6%;  Score 524;  DB 13;  Length 717;
  Best Local Similarity  40.3%;  Pred. No. 4.01e-83;
  Matches   81;  Conservative   37;  Mismatches  66;  Indels  17;  Gaps  12;

Db    293 HKPCQNGATCTNTGQGSYTCSCRPGFTGDSCEIEVNECSGSPCRNGGSCTDLENT-YSCT 351
          : || | | |  |:   |||:| ||| |  || ::: | |||| | ::|   :   ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    352 CPPGFYGRNCE--LSA-M-T-CADGP-CFNGGH-CAD---N--P-E-GGYFCQCPMGYAG 397
          |  || |  ||  :   : : | :|  | :  : |::   |  : :  :| |:||| | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    398 FNCEKKIDHCSS--NPCSNDAQCLDLVDSYLCQCPEGFTGTHCEDNIDECATYPCQNGGT 455
          :|| |:: |:   ||| |:: |: :  || | |: |||| :|| |||:|    |||||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    456 CQDGLSDYTCTCPPGYTGKNC 476
          | ||:  | | | |||:|  |
Qy    365 CVDGILSYDCLCRPGYAGQYC 385


RESULT   15
ID   Q99108      PRELIMINARY;      PRT;   832 AA.
AC   Q99108;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS DELTA PROTEIN PRECURSOR (VERSION 2).
GN   DL.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 89196890.
RA   KOPCZYNSKI C.C., ALTON A.K., FECHTEL K., KOOH P.J.,
RA   MUSKAVITCH M.A.T.;
RT   "Delta, a Drosophila neurogenic gene, is transcriptionally complex
RT   and encodes a protein related to blood coagulation factors and
RT   epidermal growth factor of vertebrates.";
RL   GENES DEV. 2:1723-1735(1988).
RN   [2]
RP   SEQUENCE OF 422-621 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D., VAESSIN H.,
RA   CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
RN   [3]
RP   PATTERN OF TRANSCRIPTION.
RX   MEDLINE; 91209246.
RA   HAENLIN M., KRAMATSCHEK B., CAMPOS-ORTEGA J.A.;
RT   "The pattern of transcription of the neurogenic gene Delta of
RT   Drosophila melanogaster.";
RL   DEVELOPMENT 110:905-914(1990).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM.
CC       DL IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL
CC       CELL LINEAGES.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN
CC       INSECTS, THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN LEVEL, IT IS
CC       CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY COMPETE FOR
CC       BINDING WITH THE NOTCH PROTEIN.
CC   -!- SIMILARITY: THE PROTEIN INCLUDES 9 EGF-LIKE REPEATS.
CC   -!- SIMILARITY: TO DROSOPHILA SERRATE PROTEIN
CC       (AC P18168), AND VERTEBRATE BLOOD COAGULATION FACTOR IX.
DR   EMBL; Y00222; G577774; -.
DR   FLYBASE; FBgn0000463; Dl.
DR   PFAM; PF00008; EGF; 8.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; TRANSMEMBRANE; EGF-LIKE DOMAIN;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26    832       DELTA PROTEIN.
FT   DOMAIN      26    595       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN     217    566       9 EGF-TYPE REPEATS.
FT   TRANSMEM   569    617       POTENTIAL.
FT   DOMAIN     618    832       INTRACELLULAR (POTENTIAL).
FT   REPEAT     217    256       EGF-LIKE 1.
FT   REPEAT     257    290       EGF-LIKE 2.
FT   REPEAT     291    330       EGF-LIKE 3.
FT   REPEAT     331    373       EGF-LIKE 4.
FT   REPEAT     374    417       EGF-LIKE 5.
FT   REPEAT     418    452       EGF-LIKE 6.
FT   REPEAT     453    490       EGF-LIKE 7.
FT   REPEAT     491    528       EGF-LIKE 8.
FT   REPEAT     529    566       EGF-LIKE 9.
FT   CARBOHYD    98     98       POTENTIAL.
FT   CARBOHYD   137    137       POTENTIAL.
```

```
FT   CARBOHYD    167    167       POTENTIAL.
FT   CARBOHYD    421    421       POTENTIAL.
FT   CARBOHYD    649    649       POTENTIAL.
FT   CONFLICT    437    438       GK -> ET (IN REF. 2).
FT   CONFLICT    443    443       A -> S (IN REF.2).
FT   CONFLICT    459    459       G -> A (IN REF. 2).
FT   CONFLICT    490    490       S -> T (IN REF. 2).
FT   CONFLICT    591    591       T -> A (IN REF. 2).
SQ   SEQUENCE   832 AA;  88813 MW;  CF9ABEC1 CRC32;

  Query Match           9.6%;  Score 521;  DB 5;  Length 832;
  Best Local Similarity  37.1%;  Pred. No. 1.80e-82;
  Matches   95;  Conservative   55;  Mismatches  77;  Indels  29;  Gaps  22;

Db    295 CTNHRPCKNGGTCFNTGEGLYTCKCAPGYSGDDCENEIYSCDADVNPCQNGGTCIDEPHT 354
          | |  |||| : |  |:   |||:|:||: |  |||:| :| :  :|| | :||   :::
Qy    183 CLN-SPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYG--SPCLNNATC-KVAQA 238

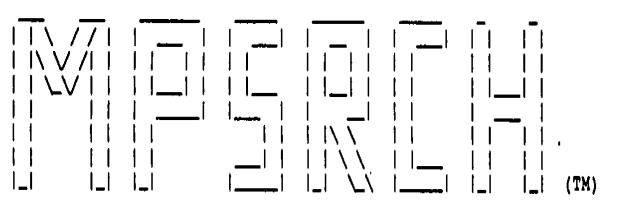
Db    355 KTGYKCHCANGWSGKMCEEKVLTCSDKPCHQG-ICRN-VR--PG-LGS-KG-Q--GYQCE 405
            ::| | :|  |  || ::  | :  |  |  | : ||  :  | :  : |  :|:|:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db    406 CPIGYSGPNCDLQLDNCSP--NPCINGGSCQP-SGK--CICPAGFSGTRCETNIDDCLGH 460
          ||: | | :|:  |: |:   ||| | | | | :|   |:|::||:|  ||||||||
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db    461 QCENGGTCIDMVNQYRCQCVPGFHGTHC--SSKVDL-CL-IRPC--AN-G-GTCL-NLNN 511
          :|:|||:|:| :  | | | ||: |  |  ::::|:      :|  :  | | |: : |:
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db    512 -DYQCTCRAGFTGKDC 526
           |: | |: ||:|  |
Qy    418 SDFTCKCHEGFSGPSC 433


Search completed: Fri May 28 09:15:05 1999
Job time : 116 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:13:09 1999; MasPar time 43.77 Seconds
916.462 Million cell updates/sec
[illegible]ar output not generated.

Title: >US-09-191-647-9
Description: (1-735) from US09191647.pep
Perfect Score: 5438
Sequence: 1 SNKNLTSFPSRIPFDTTELY.....TVHIIRQCQCEPTKSVLSEK 735

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 49.268; Variance 89.604; scale 0.550

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

[illegible]lt No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4486	82.5	601	5	Q20204	F40E10.4 PROTEIN (FRAG	0.00e+00
2	1651	30.4	1531	11	O88279	MEGF4.	0.00e+00
3	1551	28.5	1523	11	O88280	MEGF5.	0.00e+00
4	1424	26.2	739	4	O75094	MEGF5 (FRAGMENT).	9.79e-287
5	1188	21.8	530	5	Q24526	SLIT LOCUS ENCODING A	1.98e-232
6	578	10.6	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	6.53e-95
7	570	10.5	2447	13	O13149	NOTCH 2 (FRAGMENT).	3.70e-93
8	565	10.4	529	5	Q25058	FIBROPELLIN IA (FRAGME	4.60e-92
9	559	10.3	406	5	Q25059	FIBROPELLIN III (FRAGM	9.44e-91
10	540	9.9	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	1.32e-86
11	540	9.9	2470	11	O35516	CELL SURFACE PROTEIN.	1.32e-86
12	533	9.8	728	13	Q90656	TRANSMEMBRANE PROTEIN	4.42e-85
13	528	9.7	721	13	Q91902	X-DELTA-1.	5.41e-84
14	524	9.6	717	13	P87357	DELTAD TRANSMEMBRANE P	4.01e-83
15	521	9.6	832	5	Q99108	NEUROGENIC LOCUS DELTA	1.80e-82
16	522	9.6	1476	13	Q90285	PUTATIVE EXTRACELLULAR	1.09e-82
17	516	9.5	752	13	O42374	NOTCH RECEPTOR PROTEIN	2.18e-81
18	519	9.5	2531	5	O16004	NOTCH HOMOLOG.	4.88e-82
19	499	9.2	642	13	P79941	NOTCH LIGAND X-DELTA-2	1.05e-77
20	500	9.2	802	13	O57462	DELTAA.	6.38e-78
21	495	9.1	2352	5	O61240	HRNOTCH PROTEIN.	7.67e-77
22	492	9.0	615	13	O57409	DELTAB.	3.41e-76
23	489	9.0	1218	4	Q15816	TRANSMEMBRANE PROTEIN	1.51e-75
24	489	9.0	1218	4	O14902	TRANSMEMBRANE PROTEIN	1.51e-75
25	487	9.0	1218	4	O15122	JAGGED1.	4.09e-75
26	489	9.0	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	1.51e-75
27	477	8.8	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	5.83e-73
28	480	8.8	1372	5	P91526	SIMILARITY TO MULTIPLE	1.32e-73
29	471	8.7	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	1.14e-71
30	474	8.7	1219	11	Q63722	JAGGED PROTEIN.	2.57e-72
31	461	8.5	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	1.59e-69
32	463	8.5	1964	11	O35442	NOTCH4.	5.92e-70
33	444	8.2	1202	11	P97607	JAGGED2 (FRAGMENT).	6.87e-66
34	445	8.2	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	4.20e-66
35	447	8.2	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.57e-66
36	426	7.8	434	11	O55139	JAGGED2 PROTEIN (FRAGM	4.66e-62
37	422	7.8	1999	4	Q99940	NOTCH4.	3.29e-61
38	422	7.8	2003	4	O00306	NOTCH4.	3.29e-61
39	417	7.7	955	4	Q99466	NOTCH4 (FRAGMENT).	3.77e-60
40	408	7.5	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	3.02e-58
41	404	7.4	383	11	O70534	ZOG.	2.11e-57
42	404	7.4	383	11	Q62779	PREADIPOCYTE FACTOR 1.	2.11e-57
43	384	7.1	387	11	Q06007	NOTCH PROTEIN HOMOLOG	3.37e-53
44	378	7.0	156	5	Q26661	EPIDERMAL GROWTH FACTO	6.06e-52
45	380	7.0	473	5	Q25464	ADHESIVE PLAQUE MATRIX	2.31e-52

ALIGNMENTS

RESULT 1
ID Q20204 PRELIMINARY; PRT; 601 AA.
AC Q20204;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F40E10.4 PROTEIN (FRAGMENT).
GN F40E10.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA SMYE R.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z69792; E1346469; -.
DR PROSITE; PS01187; EGF_CA; 1.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 601 AA; 66669 MW; F8A72773 CRC32;

Query Match 82.5%; Score 4486; DB 5; Length 601;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAIC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 135 IKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAIC 194

```
Db     61 ETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    195 ETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYC 254

Db    121 EKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    255 EKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYC 314

Db    181 TKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    315 TKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDC 374

Db    241 LCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCD 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    375 LCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCD 434

Db    301 RQMSVGFKNPGAYLALDPLASDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    435 RQMSVGFKNPGAYLALDPLASDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVY 494

Db    361 YIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVENSGKSDQLITKGKE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    495 YIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVENSGKSDQLITKGKE 554

Db    421 MLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCSATV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    555 MLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCSATV 614

Db    481 NFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDEKRIKCDKQKFRRHHIENECRSV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    615 NFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDEKRIKCDKQKFRRHHIENECRSV 674

Db    541 DRIKIAECNGYCGGEQNCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQCEPTKSVLSE 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    675 DRIKIAECNGYCGGEQNCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQCEPTKSVLSE 734

Db    601 K 601
          |
Qy    735 K 735
```

RESULT 2
ID 088279 PRELIMINARY; PRT; 1531 AA.
AC 088279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 30.4%; Score 1651; DB 11; Length 1531;
Best Local Similarity 39.4%; Pred. No. 0.00e+00;
Matches 259; Conservative 162; Mismatches 199; Indels 37; Gaps 25;

```
Qy    748 SNKHLQALPKGIPKNVTELYLDGNQFTLVPGQ-LSTFKY-LQLVDLSNNKISSLSNSSFT 805
```

```
Qy      1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRL-YSLTKLDLSHNRLISLENNTFS 59

Db    806 NMSQLTTLILSYNALQCIPPLAFQGLRSLRLLSLHGNDVSTLQEGIFADVTSLSHLAIGA 865
          |:::|:|||:||| |:|: |||| || :||:||||||:| | :: |:::|::|:|:|:|:
Qy     60 NLTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGS 119

Db    866 NPLYCDCHLRWLSSWVKTGYKEPGIARCAGPPEMEGKLLLTTPAKKFECQGPPSLAVQAK 925
          |:||||:: |:| |:|: : |:|||| | :  ||||: : | |:: : : :|
Qy    120 NSLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATK 179

Db    926 CDPCLSSPCQNQGTCHNDPLEVYRCTCPSGYKGRNCEVSLDSCSSNPCGNGGTCHAQEGE 985
          || ||:||| | : |  :  | | |:|: | :||  :|:| ::|| | :||  ::
Qy    180 CDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQA- 238

Db    986 DAGFTCSCPSGFEGLTCGMNTDDCVKHDCVNGG-----V--C---V-D--G--IGNYTCQ 1030
          | | |  |||| | :| ||||: | |||    | |  : : : | :| |:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db   1031 CPLQYTGRACEQLVDFCSPDLNPCQHEAQCVGTPEGPRCECVPGYTGDNCSKNQDDCKDH 1090
          ||::| |: ||: :::|:  ||||:::: |:    :  | | ||:||:||  | ||||:
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db   1091 QCQNGAQCVDEINSYACLCAEGYSGQLCEIPPAPRNSCEGTE-CQ-N--G-ANCVD-QGS 1144
          :|||: ||| | || |||  ||:|| |||||      : |: || :  | ::||  | |
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db   1145 RP-VCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANITLQVSTAEDNGILLYN 1203
            | | ||:|| |:: :|| | :  :|| :  | :   : ||: : |:   |||||
Qy    418 SDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPLAS-D-GTITMTLRTTSKIGILLYY 475

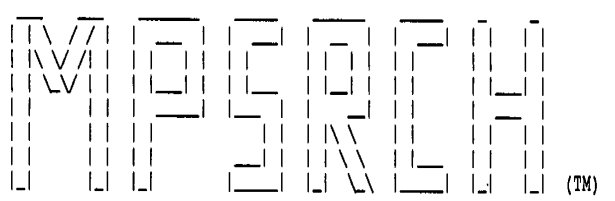
Db   1204 GDNDHIAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVTFDQMVNLSIDGGS 1263
          ||:  :: |||:|:|:: |  |::|:| :||:  :|||  | : : | :::  | ||  :
Qy    476 GDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNP 535

Db   1264 PMTMDNFGKHYTLNSEA-P-LYVGGMPVDVNSAAFRLWQILNGTSFHGCIRNLYINN--- 1318
            ::| ||   | : :   ||:||:|:: :  | |  :: |: |: ||| :: ||:
Qy    536 VQIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPI 595

Db   1319 ELQD-FTKTQMKPGVVPGCEPCRKLYCLHGICQPNA-TP-GPVCHCEAGWGGLHCDQ 1372
          :||: : :    :  :  : |  : | :| ` ||  :| | :|:|::  :| |||:
Qy    596 NLQQALENVNTEQSCSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDE 652
```

RESULT 3
ID 088280 PRELIMINARY; PRT; 1523 AA.
AC 088280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 28.5%; Score 1551; DB 11; Length 1523;
Best Local Similarity 36.7%; Pred. No. 0.00e+00;
Matches 241; Conservative 173; Mismatches 209; Indels 34; Gaps 17;



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:22:40 1999; MasPar time 53.13 Seconds
610.419 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-2
Description: (1-1525) from US09191647.pep
Perfect Score: 11299
Sequence: 1 MRGVGWQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 40.463; Variance 233.972; scale 0.173

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	7810	69.1	1534	30	W46966	Amino acid sequence o	0.00e+00
2	4159	36.8	1480	5	R25079	Drosophila SLIT prote	1.40e-293
3	1039	9.2	228	30	W46967	Amino acid sequence o	7.55e-63
4	838	7.4	183	30	W46968	Amino acid sequence o	2.62e-48
5	740	6.5	1872	36	W68510	Partial human Notch-3	2.86e-41
6	734	6.5	2321	36	W49698	Human Notch3 protein.	7.69e-41
7	666	5.9	727	21	W11719	C-Delta-1 polypeptide	5.53e-36
8	666	5.9	740	21	W00876	C-Delta-1 polypeptide	5.53e-36
9	650	5.8	1036	25	W18351	Proliferation and dif	7.61e-35
10	650	5.8	1187	25	W18352	Proliferation and dif	7.61e-35
11	655	5.8	1218	25	W18354	Proliferation and dif	3.35e-35
12	650	5.8	1218	19	W05833	Human Serrate-1 (HJ1)	7.61e-35
13	650	5.8	1218	29	W44301	Human serrate 1.	7.61e-35
14	644	5.7	520	25	W18348	Proliferation and dif	2.03e-34
15	644	5.7	702	25	W18349	Proliferation and dif	2.03e-34
16	644	5.7	723	25	W18353	Proliferation and dif	2.03e-34
17	647	5.7	1055	29	W44298	Human serrate 2 prote	1.24e-34
18	648	5.7	1193	19	W05835	Chick Serrate.	1.06e-34
19	649	5.7	1208	28	W40827	Human Jagged protein.	8.96e-35
20	647	5.7	1212	29	W44299	Human serrate 2.	1.24e-34
21	647	5.7	1257	19	W05834	Human Serrate-2 (HJ2)	1.24e-34
22	627	5.5	722	21	W11720	M-Delta-1 polypeptide	3.28e-33
23	608	5.4	612	28	W39256	Human partial mature	7.30e-32
24	608	5.4	737	28	W39257	Human membrane protei	7.30e-32
25	581	5.1	685	37	W80813	Nucleotide sequence o	5.95e-30
26	554	4.9	833	6	R28960	Delta Dll.	4.78e-28
27	507	4.5	473	17	R86869	Adhesive protein.	9.57e-25
28	494	4.4	560	12	R71294	Human glycoprotein V.	7.77e-24
29	492	4.4	660	21	W11725	H-Delta-1 polypeptide	1.07e-23
30	486	4.3	1091	27	W41641	Sequence used in dete	2.82e-23
31	464	4.1	1404	7	R38304	Sequence of a serrate	9.61e-22
32	432	3.8	157	21	W11730	H-Delta-1 polypeptide	1.59e-19
33	428	3.8	383	10	R56166	Neuroendocrine tumor	3.01e-19
34	405	3.6	385	10	R56167	Neuroendocrine tumor	1.16e-17
35	401	3.5	196	5	R29102	Drosophila SLIT prote	2.19e-17
36	394	3.5	345	23	W09405	Pineal gland specific	6.61e-17
37	390	3.5	353	1	R05160	Sequence of human bon	1.24e-16
38	371	3.3	2189	1	R05222	Antigen GX5401FL enco	2.48e-15
39	364	3.2	186	8	R42264	Decorin sequence PT-7	7.42e-15
40	366	3.2	234	8	R42265	Decorin sequence PT-7	5.43e-15
41	366	3.2	280	8	R42266	Decorin sequence PT-7	5.43e-15
42	366	3.2	305	8	R42267	Decorin sequence PT-7	5.43e-15
43	366	3.2	331	8	R42260	Mature decorin PT-65.	5.43e-15
44	366	3.2	342	17	R89439	Human recombinant dec	5.43e-15
45	366	3.2	1388	18	R89471	Collagen/decorin fusi	5.43e-15

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 69.1%; Score 7810; DB 30; Length 1534;
Best Local Similarity 65.4%; Pred. No. 0.00e+00;
Matches 986; Conservative 281; Mismatches 233; Indels 8; Gaps 5;

Db 28 rlgasacpalctctgttvdchgtglqaipknipmterlelngnnitrihkndfaglkql 87
Qy 22 KVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHL 81

Db 88 rvlqlmenqigavergafddmkelerlrlnrnqlhmlpellfqnnqalsrldlsenaiqa 147

```
Qy      82 RVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQA 141

Db     148 iprkafrgatdlknlrldknqiscieegafralrglevltlnnnnittipvssfnhmpkl 207
           ||||||||| |:|||:|| |||||||:|||||| |||||||||||| ::|:|||||||
Qy     142 IPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKL 201

Db     208 rtfrlhsnhlfcdchlawlsqwlrqrptiglftqcsgpaslrglnvaevqksefscsgqg 267
           ||||||||:|:|||||||||:||| || :||:||| ||: ||| |||||| || || :
Qy     202 RTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEE 261

Db     268 eagrvptctlssg-scpamctcsngivdcrgkgltaipanlpetmteirlelngiksipp 326
           |: :      |   ||| ||||| |||||||||| ||:|||||:|||||| | || |||
Qy     262 EGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPP 321

Db     327 gafspyrklrridlsnnqiaeiapdafqglrslnslvlygnkitdlprgvfgglytlqll 386
           ||||||:|||||||||||||:|:|||||||||||||||||||:||:::| ||::|||
Qy     322 GAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLL 381

Db     387 llnankincirpdafqdlqnlslllslydnkiqslakgtftslraiqtlhlaqnpficdcn 446
           |||||||||:| ||||||:||:|||||||:|::|||||::|||||:||||||||||:
Qy     382 LLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCH 441

Db     447 lkwladflrtnpietsgarcasprrlankrigqikskkfrcsakeqyfipgtedyqlnse 506
           ||||||:|:||||||||||:|||||||||||||||||: |:|    :    |:::
Qy     442 LKWLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDY--R-S---KLSGD 495

Db     507 cnsdvvcphkcrceanvvecsslkltkiperipqstaelrlnnneisileatgmfkklth 566
           | :|: || |||||:  |:||: || ||||:||| ||||||||||:::|||||:|||| :
Qy     496 CFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQ 555

Db     567 lkkinlsnnkvseiedgafegaasvselhltanqlesirsgmfrgldglrtlmlrnnris 626
           |:|||:||||:::||:||||||::|:|: ||:|:||:::  ||:||:::|:|||||:|||:
Qy     556 LRKINFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRIT 615

Db     627 cihndsftglrnvrllslydnqittvspgafdtlqslstlnllanpfncncqlawlggwl 686
           |: |||| || :|||||||||||||:|||||||:|||||||||||||||| ||||| ||
Qy     616 CVGNDSFIGLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWL 675

Db     687 rkrkivtgnprcqnpdflrqiplqdvafpdfrceegqeeggclprpqcpqecacldtvvr 746
           ||::|||||||||:| ||::||:|||||: || |::| :: :| | ::|| ||:|||||||
Qy     676 RKKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVR 735

Db     747 csnkhlralpkgipknvtelyldgnqftlvpgqlstfkylqlvdlsnnkisslsnssftn 806
           |||| |: |||||::|||||||||||||| :|| :| | |:|||||:||:||| ||:|
Qy     736 CSNKGLKVLPKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQSFSN 795

Db     807 msqlttlilsynalqcipplafqglrslrllslhgndistlqegifadvtslshlaigan 866
           |:|| ||||||| |:|||| :|:||:|||||||||||||| : || | |:::|||||||||
Qy     796 MTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGAN 855

Db     867 plycdchlrwlsswvktgykepgiarcagpqdmegkllltttpakkfecqgpptlavqakc 926
           ||||||:::||| |||: |||||||||||| :|  |||||||:||| ||||  : : |||
Qy     856 PLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKC 915

Db     927 dlclsspcqnqgtchndplevyrcacpsgykgrdcevslnscssgpcenggtchaqeged 986
           : |||:|| |:|||::||:: |||:|| |:||:||:|::::| | || :|||||  |||:
Qy     916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db     987 apftcscptgfegptcgvntddcvdhacanggvcvdgvgnytcqcplqyegkaceqlvdl 1046
             | | |: |||| | || ||| |: | | : ||||: |||| || :| |  ||: :|:
Qy     976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    1047 cspdlnpcqheaqcvgtpdgprcecmpgyagdncsenqddcrdhrcqngaqcmdevnsys 1106
           |: |||||||:: |: || | :|:| ||| |::|  : |||:|::| |||:| | ||:|:
Qy    1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    1107 clcaegysgqlceipphlpapk-spcegtecqngancvdqgnrpvcqclpgfggpecekl 1165
           |:|:||||| :||::| :  |: |||:  :||||| |: : | |:|||||||: |  ||||
Qy    1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155

Db    1166 lsvnfvdrdtylqftdlqnwpranitlqvstaedngillyngdndhiavelyqghvrvsy 1225
           :||||:::::|||:     :|::|||||::| ||:|||||:||:||||||||:|:|| ||
Qy    1156 VSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASY 1215

Db    1226 dpgsypssaiysaetindgqfhtvelvafdqmvnlsidggspmtmdnfgkhytlnseapl 1285
           | || |:||||| |||||| || |||:|:|| ::||:|||:|: : |::|: ||| ::||
Qy    1216 DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPL 1275

Db    1286 yvggmpvdvnsaafrlwqilngtgfhgcirnlyinnelqdftktqmkpgvvpgcepcrkl 1345
           ||||||  | |::|    |||:||||||||||:||||| |  |  |::|||||:|
Qy    1276 YVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKK 1335

Db    1346 yclhgicqpnatpgpmchceagwvglhcdqpadgpchghkcvhgqcvpldalsyscqcqd 1405
           | || |||:: :|  | |: ||:|  ||| :: || |:||||| |:|::|:|||| | :
Qy    1336 VCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLE 1395

Db    1406 gysgalcnqagalaepcrglqclhghcqasgtkgahcvcdpgfsgelceqesecrgdpvr 1465
           | :| ||::  | :||::: | || |: ||  : | | :|::|: |::|  |||: :|
Qy    1396 GHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYCECSSGYTGDSCDREISCRGERIR 1455

Db    1466 dfhqvqrgyaicqttrplswvecrgscpgqgccqglrlkrrkftfecsdgtsfaeevekp 1525
           |: | |:||| |||: :|::||||:|:|  ||  || |||::|||:||:|| :||||
Qy    1456 DYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKV 1515

Db    1526 tkcgcalc 1533
             ||||: |
Qy    1516 VKCGCTRC 1523


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
```

```
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match           36.8%;  Score 4159;  DB 5;  Length 1480;
  Best Local Similarity  43.9%;  Pred. No. 1.40e-293;
  Matches   592;  Conservative  335;  Mismatches 370;  Indels  52;  Gaps  33;

Db   105 lelqgnnltviyetdfqrltklrmlqltdnqihtiernsfqdlvslerldisnnvittvg 164
Qy    84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Db   165 rrvfkgaqslrslqldnnqitcldehafkglveleiltvnnnnltslphnifggvgrlra 224
Qy   144 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 203

Db   225 lrlsdnpfacdchlswlsrflrsatrlapytrcqspsqlkgqnvadlhdqefkcsglte- 283
Qy   204 FRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEEG 263

Db   284 hap-mecg-aenscphpcrcadgivdcreksltsvpvtlpddttdvrleqnfitelppks 341
Qy   264 HQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGA 323

Db   342 fssfrrlrridlsnnnisriahdalsglkqlttlvlygnkikdlpsgvfkglgslrllll 401
Qy   324 FSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLL 383

Db   402 naneiscirkdafrdlhslsllslydnniqsvangtfdamksmktvhlaknpficdcnlr 461
Qy   384 NANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLK 443

Db   462 wvadylhknpietsgarcespkrmhrrriesvreekfkcs-wgelrmklsgecrmdsdcp 520
Qy   444 WLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACP 503

Db   521 amchcegttvdctgrrlkeiprdiplhttelllndnelgrissdglfgrlphlvklelkr 580
Qy   504 EKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFSN 563

Db   581 nqltgiepnafegashiqe--l---ql-g-enki-k--e------i-snkm-------fl 616
Qy   564 NKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRITCVGNDSFI 623

Db   617 glhqlktlnlydnqiscvmpgsfehlnsltslnlasnpfncnchlawfaecvrkkslngg 676
Qy   624 GLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKKRIVTG 683

Db   677 aarsgapskvrdlqikdlphsefkcssense-gclgdgycppsctctgtvvacsrnqlke 735
Qy   684 NPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRCSNKGLKV 743

Db   736 iprgipaetsqvylesneieqihyerirhlrsltrldlsnnqitilsnytfanltklsrl 795
Qy   744 LPKGIPRDVTELYLDGNQFTLVPKE-LSNYKHLTLIDLSNNRISTLSNQSFSNMTQLLTL 802

Db   796 iisynklqclqrhalsglnnlrvvslhgnrismlpeasfedlkslthialgsnplycdcg 855
Qy   803 ILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLYCDCN 862

Db   856 lkwfsdwikldyvepgiarcaepeqmkdklilstpsssfvcrgrvrndilakcnacfeqp 915
Qy   863 MQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNP 922

Db   916 cqnqaqcvalpqreyqclcqpgyhgkhcefmidacygnpcrnnatctvle--egrfrcgc 973
Qy   923 CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCIC 982

Db   974 apgytgarcetniddclgeikcqnnatcidgvesykcecqpgfsgefcdtkiqfcspefn 1033
Qy   983 ADGFEGENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLN 1041

Db  1034 pcangakcmdhfthyscdcqagfhgtnctdniddcqnhmcqnggtcvdgindyqcrcpdd 1093
Qy  1042 PCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEG 1101

Db  1094 ytgkyceghnmismmypqtspcqnheckhgvcfqpnaqgsdylcrchpgytgkwceylts 1153
Qy  1102 YSGLFCE-FSP-PMVLPRTSPCDNFDCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVS 1157

Db  1154 isfvhnnsfveleplrtrpeanvtivfssaeqngilmydgqdahlavelfngrirvsydv 1213
Qy  1158 VNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASYDT 1217

Db  1214 gnhpvstmysfemvadgkyqavellaikknftlrvdrglarsiinegsndylklttpmfl 1273
Qy  1218 GSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYV 1277

Db  1274 gglpvdpaqqayknwqirnltsfkgcmkevwinhklvdfgnaqrqqkitpgcal-lege- 1331
Qy  1278 GGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVC 1337

Db  1332 qqee-eddeqd-fm-d--et--phikeepv-dpclenkcrrgsrcvpnsnardgyqckck 1383
Qy  1338 AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCL 1394

Db  1384 hgqrgrycdqgegstep-ptvt-aastcr 1410
Qy  1395 EGHGGVLCDEEEDLFNPCQAIKCKHGKCR 1423


RESULT   3
ID   W46967 standard; Protein; 228 AA.
AC   W46967;
DT   06-JUL-1998  (first entry)
DE   Amino acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody.
OS   Mus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   N-PSDB; V16966.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Page 35; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   228 AA;
```

```
  Query Match           9.2%;  Score 1039;  DB 30;  Length 228;
  Best Local Similarity  58.7%;  Pred. No. 7.55e-63;
  Matches  132;  Conservative   43;  Mismatches  50;  Indels   0;  Gaps   0;

Db      2 edngillyngdndhiavelyqghvrvsydpgsypssaiysaetindgqfhtvelvtfdqm 61
          ||:|||||:||:||||||||:|:|| ||| || |:|||| |||||| || |||:::||
Qy   1188 EDSGILLIYKGDKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQS 1247

Db     62 vnlsidggspmtmdnfgkhytlnseaplyvggmpvdvnsaafrlwqilngtsfhgcirnl 121
          ::||:|||:|: : |::|: ||| ::|||||||   | |::|     ||||||||||||
Qy   1248 LSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNL 1307

Db    122 yinnelqdftktqmkpgvvpgcepcrklyclhgicqpnatpgpvchceagwgglhcdqpv 181
          |||:||||| |  |  |::|||||||:|  | || |||:: :|  | |: || |  |||
Qy   1308 YINSELQDFQKVPMQTGILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRT 1367

Db    182 dgpchghkcvhgkcvpldalayscqcqdgysgalcnqvgavaepc 226
          : || |:||||| |:|::|::||| | :| :| ||::   : :||
Qy   1368 NDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPC 1412


RESULT   4
ID   W46968 standard; Protein; 183 AA.
AC   W46968;
DT   06-JUL-1998  (first entry)
DE   Amino acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody.
OS   Rattus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   N-PSDB; V16967.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Pages 37-38; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   183 AA;

  Query Match           7.4%;  Score 838;  DB 30;  Length 183;
  Best Local Similarity  59.0%;  Pred. No. 2.62e-48;
  Matches  108;  Conservative   33;  Mismatches  42;  Indels   0;  Gaps   0;

Db      1 ghvrvsydpgsypssaiysaetindgqfhtvelvtfdqmvnlsidggspmtmdnfgkhyt 60
          |:|| ||| || |:||||| |||||| || |||:::|| ::||:|||:|: : |::|: |
Qy   1209 GRVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQST 1268

Db     61 lnseaplyvggmpvdvnsaafrlwqilngtsfhgcirnlyinnelqdftktqmkpgvvpg 120
          || ::||||||||   | |::|     ||||||||||||||:||||| |  |  |::||
Qy   1269 LNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPG 1328

Db    121 cepcrklyclhgicqpnatpgpvchceagwgglhcdqpvdgpchghkcvhgkcvpldala 180
          ||||:|  | || |||:: :|  | |: || |  |||  : || |:||||| |:|::|::
Qy   1329 CEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFS 1388

Db    181 ysc 183
          |||
Qy   1389 YSC 1391


RESULT   5
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 328
FT                   /note= "encoded by NAN"
FT   Misc_difference 401
FT                   /note= "encoded by GNN"
FT   Misc_difference 403
FT                   /note= "encoded by GNC"
FT   Misc_difference 406
FT                   /note= "encoded by GNN"
FT   Misc_difference 409
FT                   /note= "encoded by NNT"
FT   Misc_difference 420
FT                   /note= "encoded by GNC"
FT   Misc_difference 706
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
SQ   Sequence   1872 AA;

  Query Match           6.5%;  Score 740;  DB 36;  Length 1872;
  Best Local Similarity  31.4%;  Pred. No. 2.86e-41;
  Matches  169;  Conservative   99;  Mismatches 229;  Indels  42;  Gaps  36;

Db     94 decrvgepcrhggtclntpgsf-rcqcpagytgplcenpavpcapspcrnggtcr-qsgd 151
          : | :::||:: ||| : |  | || || |: |  |: |  :| ::||::||||:   |:
Qy    916 NPC-LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    152 lt-ydcaclpgfeqqncevnvddcpghrclnggtcvdgvntyncqcppewtgqfctedvd 210
             : | |  ||||:|||||||||  : | | :|||||:| | | |||| ||::| | :|
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    211 ecqlqpnachnggtcfntlgghscvcvngwtgescsqniddcatavcfhgatchdrvasf 270
          |  : |:|:: : |: |  |  | |  |  || |  ::|||    | :|| | | | ::
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    271 ycacpmgktgllchlddacv---snpcredaicdtnp---vn-graictcppgftggacd 323
          | || | :||:| :  : |   ::|| ::  |:  :   |    :|| | ||: |  |:
```

```
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153

Db    324 qdvdxcsiganpcehlgrc-vntqgsflcqcgrgy-tgpr-cetdvneclsgpcrnqa-t 379
Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRA 1213

Db    380 cldrigqftc-ic-magft-gtycxvxi-dxcqx-spcvnggvckxrvn-gfsct-c-ps 431
Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDS 1273

Db    432 g-fsgstc-qldvdecastpcrngakcvd-qpdgyecrcaegfegtlcdrnv-ddcsp-d 486
Qy   1274 PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCH 1333

Db    487 p--chhgrc-vdgiasfscacapgytgtrcesqvde-crsqpcrhggkcldlvd-kylcr 541
Qy   1334 KKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 1392

Db    542 cpsgttgvncevnid--d-casnpctfgvcr-dginr-ydcvcqpgftgplcnveinec 595
Qy   1393 CLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449


RESULT    6
ID   W49698 standard; Protein; 2321 AA.
AC   W49698;
DT   21-DEC-1998  (first entry)
DE   Human Notch3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy.
OS   Homo sapiens.
PN   FR2751986-A1.
PD   06-FEB-1998.
PF   16-APR-1997; 004680.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133138/13.
DR   N-PSDB; V57001.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1.1-1.8; 45pp; French.
CC   This sequence represents the human Notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of
CC   the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC   and leukoencephalopathy (CADASIL) type.  Blocking expression of a
CC   mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC   gene or protein can be used to treat CADASIL or related disorders.
SQ   Sequence   2321 AA;

  Query Match           6.5%;  Score 734;  DB 36;  Length 2321;
  Best Local Similarity  31.5%;  Pred. No. 7.69e-41;
  Matches  170;  Conservative  100;  Mismatches 227;  Indels  42;  Gaps  36;

Db    160 decrvgepcrhggtclntpgsf-rcqcpagytgplcenpavpcapspcrnggtcr-qsgd 217
Qy    916 NPC-LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    218 lt-ydcaclpgfegqncevnvddcpghrclnggtcvdgvntyncqcppewtgqfctedvd 276
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    277 ecqlqpnachnggtcfntlgghscvcvngwtgescsqniddcatavcfhgatchdrvasf 336
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    337 ycacpmgktgllchlddacv---snpchedaicdtnp---vn-graictcppgftggacd 389
```

```
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153

Db    390 qdvdecsiganpcehlgrc-vntqgsflcqcgrgy-tgpr-cetdvneclsgpcrnqa-t 445
Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRA 1213

Db    446 cldrigqftc-ic-magft-gtycevdi-decqs-spcvnggvckdrvn-gfsct-c-ps 497
Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDS 1273

Db    498 g-fsgstc-qldvdecastpcrngakcvd-qpdgyecrcaegfegtlcdrnv-ddcsp-d 552
Qy   1274 PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCH 1333

Db    553 p--chhgrc-vdgiasfscacapgytgtrcesqvde-crsqpcrhggkcldlvd-kylcr 607
Qy   1334 KKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 1392

Db    608 cpsgttgvncevnid--d-casnpctfgvcr-dginr-ydcvcqpgftgplcnveinec 661
Qy   1393 CLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449


RESULT    7
ID   W11719 standard; Protein; 727 AA.
AC   W11719;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide.
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          184..228
FT                   /label= DSL
FT   domain          229..261
FT                   /label= EGF1
FT   domain          262..292
FT                   /label= EGF2
FT   domain          293..332
FT                   /label= EGF3
FT   domain          333..370
FT                   /label= EGF4
FT   domain          371..409
FT                   /label= EGF5
FT   domain          410..447
FT                   /label= EGF6
FT   domain          448..485
FT                   /label= EGF7
FT   domain          486..523
FT                   /label= EGF8
FT   domain          524..534
FT                   /label= EGF9
FT   domain          555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58897.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W11719) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
```

```
CC  C-Delta-1 correlates with onset of neurogenesis. The C-delta-1
CC  amino acid sequence was deduced from a cDNA clone (T58897) obtd.
CC  from chick stage 4-6 embryos. An alternatively spliced variant
CC  (W00876) was also isolated, and mouse (W11720) and human (W11721-
CC  38) Delta-1 polypeptides have been identified. Delta-1 proteins
CC  can be used to treat or prevent disorders characterised by
CC  increased Notch activity, such as cervical, breast, lung or colon
CC  cancer, melanoma or seminoma, and nervous system disorders or to
CC  promote tissue regeneration and repair.
SQ  Sequence  727 AA;

  Query Match           5.9%;  Score 666;  DB 21;  Length 727;
  Best Local Similarity  39.5%;  Pred. No. 5.53e-36;
  Matches   94;  Conservative   50;  Mismatches  81;  Indels  13;  Gaps   9;

Db   303 kpckngatctntgqgsytcscrpgytgssceieinecdanpcknggsc--td-lensysc 359
Qy   921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db   360 tcppgfygkncelsamtcadgpcfnggrctdnpdggyscrcplgysgfncekkidycs-s 418
Qy   981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db   419 -spcangaqcvdlgnsyicqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcp 477
Qy  1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db   478 pgyngknc--stp-v-sr---cehnpchngatchersnryvcecargygglncqfllp 528
Qy  1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT   8
ID  W00876 standard; Protein; 740 AA.
AC  W00876;
DT  28-APR-1997  (first entry)
DE  C-Delta-1 polypeptide (alternatively spliced variant).
KW  C-Delta-1; cell proliferation; nervous system disorder;
KW  tissue regeneration; Notch; cervix cancer; breast cancer;
KW  lung cancer; colon cancer; melanoma; seminoma;
KW  neurogenesis; therapy.
OS  Gallus sp.
FH  Key             Location/Qualifiers
FT  domain          184..228
FT                  /label= DSL
FT  domain          229..261
FT                  /label= EGF1
FT  domain          262..292
FT                  /label= EGF2
FT  domain          293..332
FT                  /label= EGF3
FT  domain          333..370
FT                  /label= EGF4
FT  domain          371..409
FT                  /label= EGF5
FT  domain          410..447
FT                  /label= EGF6
FT  domain          448..485
FT                  /label= EGF7
FT  domain          486..523
FT                  /label= EGF8
FT  domain          524..534
FT                  /label= EGF9
FT  domain          555..579
FT                  /label= TM
FT                  /note= "transmembrane domain"
PN  WO9701571-A1.
PD  16-JAN-1997.
PF  28-JUN-1996; U11178.
PR  28-JUN-1995; US-000589.
PA  (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA  (UYYA ) UNIV YALE.
PI  Artavanis-Tsakonas S, Gray GE, Henrique D, Ish-Horowicz D;
PI  Lewis J;
DR  WPI; 97-100159/09.
DR  N-PSDB; T58898.
PT  New vertebrate Delta protein, DNA and antibodies - for treating and
PT  preventing cancer, nervous system disorders and for tissue
PT  regeneration
PS  Disclosure; Fig 2; 135pp; English.
CC  C-delta-1 polypeptide (W00876) is the chick homologue of Drosophila
CC  Delta, a protein that binds to Notch protein. Expression of
CC  C-Delta-1 correlates with onset of neurogenesis. The C-delta-1
CC  amino acid sequence was deduced from a cDNA clone (T58898) obtd.
CC  from chick stage 4-6 embryos. A shorter version (W58877) of
CC  C-Delta-1, lacking the 12 C-terminal amino acids of the longer
CC  version, was also isolated, and mouse (W11720) and human (W11721-
CC  38) Delta-1 polypeptides have been identified. Delta-1 proteins
CC  can be used to treat or prevent disorders characterised by
CC  increased Notch activity, such as cervical, breast, lung or colon
CC  cancer, melanoma or seminoma, and nervous system disorders or to
CC  promote tissue regeneration and repair.
SQ  Sequence  740 AA;

  Query Match           5.9%;  Score 666;  DB 21;  Length 740;
  Best Local Similarity  39.5%;  Pred. No. 5.53e-36;
  Matches   94;  Conservative   50;  Mismatches  81;  Indels  13;  Gaps   9;

Db   303 kpckngatctntgqgsytcscrpgytgssceieinecdanpcknggsc--td-lensysc 359
Qy   921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db   360 tcppgfygkncelsamtcadgpcfnggrctdnpdggyscrcplgysgfncekkidycs-s 418
Qy   981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db   419 -spcangaqcvdlgnsyicqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcp 477
Qy  1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db   478 pgyngknc--stp-v-sr---cehnpchngatchersnryvcecargygglncqfllp 528
Qy  1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT   9
ID  W18351 standard; protein; 1036 AA.
AC  W18351;
DT  11-FEB-1998  (first entry)
DE  Proliferation and differentiation suppression polypeptide.
KW  Proliferation; differentiation; suppression; human; delta-1;
KW  serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW  immunosuppression.
OS  Homo sapiens.
PN  WO9719172-A1.
PD  29-MAY-1997.
PF  15-NOV-1996; J03356.
PR  30-NOV-1995; JP-311811.
PR  17-NOV-1995; JP-299611.
PA  (ASAH ) ASAHI KASEI KOGYO KK.
PI  Itoh A, Sakano S;
DR  WPI; 97-298110/27.
PT  Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT  proliferation and differentiation of undifferentiated human blood
PT  cells
PS  Claim 5; Page 66-71; 114pp; Japanese.
CC  The present sequence represents a polypeptide which suppresses
CC  proliferation and differentiation of undifferentiated cells such
CC  as neurons and blood cells. The polypeptide may be used for the
CC  prevention and control of disorders involving undifferentiated
CC  cells, such as leukaemia and malignant tumours, and improvement of
CC  blood formation, e.g. after immunosuppression.
SQ  Sequence  1036 AA;

  Query Match           5.8%;  Score 650;  DB 25;  Length 1036;
  Best Local Similarity  36.7%;  Pred. No. 7.61e-35;
```

```
  Matches   88;  Conservative   53;  Mismatches  85;  Indels  14;  Gaps   7;

Db    307 haclsdpchnrgscketslgf-ececspgwtgptcstniddcspnncshggtcq--d-lv 362
          ::|||:|| | |:|:  :: |  | |: |  |  |   | | :| | |||||:  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    363 ngfkcvcppqwtgktcqldaneceakpcvnakscknliasyycdclpgwmgqncdinind 422
          :|| |:|:    |  |::: ::|| : | |  :| : | :| | | |   |: |: :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    423 c---lgqcqndascrdlvngyrcicppgyagdhcerdidecasnpclngghcqneinrfq 479
          |   |  ||:|: |    :|::| | ||| |:||: |:|:|  | | ||:|| : :| :
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    480 clcptgfsgnlcqld---i---dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 532
          |:|| |:|| :|::    :     |:   |||||||  | :: :| |   |:| :|  |
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155


RESULT  10
ID   W18352 standard; protein; 1187 AA.
AC   W18352;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 6; Page 71-76; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
     Sequence   1187 AA;

  Query Match           5.8%;  Score 650;  DB 25;  Length 1187;
  Best Local Similarity  36.7%;  Pred. No. 7.61e-35;
  Matches   88;  Conservative   53;  Mismatches  85;  Indels  14;  Gaps   7;

Db    307 haclsdpchnrgscketslgf-ececspgwtgptcstniddcspnncshggtcq--d-lv 362
          ::|||:|| | |:|:  :: |  | |: |  |  |   | | :| | |||||:  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    363 ngfkcvcppqwtgktcqldaneceakpcvnakscknliasyycdclpgwmgqncdinind 422
          :|| |:|:    |  |::: ::|| : | |  :| : | :| | | |   |: |: :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    423 c---lgqcqndascrdlvngyrcicppgyagdhcerdidecasnpclngghcqneinrfq 479
          |   |  ||:|: |    :|::| | ||| |:||: |:|:|  | | ||:|| : :| :
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    480 clcptgfsgnlcqld---i---dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 532
          |:|| |:|| :|::    :     |:   |||||||  | :: :| |   |:| :|  |
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155


RESULT  11
ID   W18354 standard; protein; 1218 AA.
AC   W18354;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Signal
FT   Protein         32..1218
FT                   /label= Differentiation_suppression_protein
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
DR   N-PSDB; T70175.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 15; Page 83-91; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1218 AA;

  Query Match           5.8%;  Score 655;  DB 25;  Length 1218;
  Best Local Similarity  29.8%;  Pred. No. 3.35e-35;
  Matches  164;  Conservative  111;  Mismatches 226;  Indels  50;  Gaps  40;

Db    305 pclnggtcsntgpdkyqcscpegysgpnceiaehaclsdpchnrgsc--ketsl-gfece 361
          || | |||::   | |:|:|| |: | :|::: |||:|:|| : |:|  ||    || |
Qy    922 PCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCI 981

Db    362 cspgwtgptcstniddcspnncshggtcqdlvngfkcvcppqwtgktcqldanqc-ea-k 419
          |: |  |  |  |:|||  |:| : :|| | :| : |:|||: ||  |:   : | :  :
Qy    982 CADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLN 1041

Db    420 pcvnakscknlia-syycdclpgwmgqncdinindclgq-cqndascrdlvngyrcicpp 477
          || :  |  | : :: ||| || :|::|||:::||    | | | | | |||| ||||
Qy   1042 PCQHDSKCI-LTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPE 1100

Db    478 gyagdhce--rdidecasnpclngghcqneinrfqclcptgfsgnlcqldidycepnpcq 535
          ||:|  ||    :    ::|| :  |||     ||:    :: :||     | |   |
Qy   1101 GYSGLFCEFSPPMVLPRTSPC-DNFDCQNG-A--QCIVR--INEPICQ-----CLPG-YQ 1148

Db    536 ngaqcynrasdyfckcpedyegkncshlkdhcrtt-pcevidsctvamasndtpe-gvry 593
          |  | :  |  | :  | |    : :: :  .|      :  : :  |    :|
Qy   1149 -GEKCEKLVSVNFIN-KESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVEL 1206

Db    594 issnvcgphgkcksqsggkftcdc-nkg-ftgtychenindsesnpcrnggt-cidgvns 650
            : | ::       ::: :: :  | | |        :::| | :   |    |  ::
Qy   1207 YRGRVRASYDTGSHPASAIYSVETINDGNFHIVE-LLALDQSLSLSVDGGNPKIITNLSK 1265

Db    651 ykcicsdg--wegayce-tnindcsqnpchnggtcrdlvndfycdck-ngw-k-gk-tch 703
            :  |:    |:   :|:    | | :|| : :  : ::| :       |  : |
Qy   1266 QSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGI 1325

Db    704 srdsq-cdeatcnnggtcydegda-fkcmcpggwegttcniarnssclpnpchnggtcv- 760
            : |    | :| ||   ::| | | |  || |  |:   | :|| | | :| ||:
Qy   1326 LPGCEPCHKKVCAHG-TCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLP 1383

Db    761 vngesftcvckegwegpicaqntn--d-csphpcynsgtc-vdg-dnwyrcecapgfagp 815
          :|: |::| | ||  | :| :: :  : | :  | : | | : |    | |||::|::|
Qy   1384 INAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKC-KHGKCRLSGLGQPY-CECSSGYTGD 1441

Db    816 dcrininecqs 826
          |  :|: |::
```

```
Qy   1442 SCDREIS-CRG 1451


RESULT  12
ID   W05833 standard; Protein; 1218 AA.
AC   W05833;
DT   28-JAN-1997  (first entry)
DE   Human Serrate-1 (HJ1).
KW   Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
KW   cell fate; central nervous system; cancer; tissue repair; therapy;
KW   diagnosis; antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1..1067
FT                   /label= Extracellular_domain
FT   peptide         14..29
FT                   /label= Sig_peptide
FT   domain          185..229
FT                   /label= DSL
FT                   /note= "region of homology with Drosophila Delta
FT                   and Serrate, predicted to mediate binding
FT                   with Notch"
FT   domain          234..896
FT                   /label= ELR
FT                   /note= "epidermal growth factor-like repeat domain"
FT   region          234..264
FT                   /label= ELR1
FT   region          265..299
FT                   /label= ELR2
FT   region          300..339
FT                   /label= ELR3
FT   region          340..377
FT                   /label= ELR4
FT   region          378..415
FT                   /label= ELR5
FT   region          416..453
FT                   /label= ELR6
FT   region          454..490
FT                   /label= ELR7
FT   region          491..528
FT                   /label= ELR8
FT   region          529..566
FT                   /label= ELR9
FT   region          567..598
FT                   /label= Partial_ELR
FT   region          599..632
FT                   /label= Partial_ELR
FT   region          633..670
FT                   /label= ELR10
FT   region          671..708
FT                   /label= ELR11
FT   region          709..747
FT                   /label= ELR12
FT   region          748..785
FT                   /label= ELR13
FT   region          786..823
FT                   /label= ELR14
FT   region          824..862
FT                   /label= ELR15
FT   region          863..879
FT                   /label= Partial_ELR
FT   region          880..896
FT                   /label= Partial_ELR
FT   domain          1068..1089
FT                   /label= Transmembrane_domain
FT   domain          1090..1218
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique DMP,  Ish-Horowicz D;
PI   Lewis JH, Mann RS,  Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40090.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Claim 4; Page 95-98; 161pp; English.
CC   Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC   for the zygotic neurogenic locus Notch, and are believed to play a
CC   major role in determining cell fates (differentiation) in the
CC   central nervous system.  Their amino acid sequences were deduced
CC   from cDNA clones (see also T40090-91) isolated from human foetal
CC   brain cDNA libraries.  The proteins, antibodies raised to them,
CC   and encoding nucleic acids can be used in the detection of
CC   Serrate sequences and in the treatment of disorders of cell fate
CC   or differentiation, partic. cancer, nervous system disorders
CC   and in tissue repair or regeneration.
SQ   Sequence   1218 AA;

  Query Match           5.8%;  Score 650;  DB 19;  Length 1218;
  Best Local Similarity  36.7%;  Pred. No. 7.61e-35;
  Matches   88;  Conservative   53;  Mismatches  85;  Indels  14;  Gaps   7;

Db    338 haclsdpchnrgscketslgf-ececspgwtgptcstniddcspnncshggtcq--d-lv 393
          ::|||:|| | |:|:  :: |  | |: |  |  |   |  | :| | |||||:  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    394 ngfkcvcppqwtgktcqldaneceakpcvnakscknliasyycdclpgwmgqncdinind 453
          :|| |:|:    |  |::: ::|| : | |  :| : | :| | | |   |: |: :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    454 c---lgqcqndascrdlvngyrcicppgyagdhcerdidecasnpclngghcqneinrfq 510
          |   |  ||:|: |    :|::| | ||| |:||: |:|:|  | | ||:|| : :| :
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    511 clcptgfsgnlcqld---i---dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 563
          |:|| |:|| :|::    :      |:  ||||||  | :: :| |   |:| :|  |
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155


RESULT  13
ID   W44301 standard; Protein; 1218 AA.
AC   W44301;
DT   19-JUN-1998  (first entry)
DE   Human serrate 1.
KW   Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW   leukaemia; endothelial cell; tumour.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Signal
FT   Protein         32..1218
FT                   /label= Serrate-1
PN   WO9802458-A1.
PD   22-JAN-1998.
PF   11-JUL-1997; J02414.
PR   14-MAY-1997; JP-124063.
PR   16-JUL-1996; JP-186220.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A, Sakano S;
DR   WPI; 98-110528/10.
DR   N-PSDB; V15201.
PT   Human serrate-2 gene expression products  - used to regulate stem
PT   cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS   Disclosure; Page 77-86; 103pp; Japanese.
CC   The present sequence represents human serrate 1, from the present
CC   invention which describes human serrate 2. The present invention also
CC   describes a method for the preparation of the polypeptides, and
CC   antibodies binding to the polypeptide and its fragments. The polypeptide
CC   and its fragments expressed by the serrate-2-gene can be used to inhibit
CC   stem (especially blood stem) cell differentiation and to inhibit
CC   endothelial cell growth. They may be incorporated in a cell culture
CC   media for culturing undifferentiated stem cells. They can also be used
```

```
CC   for treatment of neoplasms such as leukaemia. The antibodies can be used
CC   for the diagnosis of malignant tumours.
SQ   Sequence   1218 AA;

  Query Match           5.8%;  Score 650;  DB 29;  Length 1218;
  Best Local Similarity  36.7%;  Pred. No. 7.61e-35;
  Matches   88;  Conservative   53;  Mismatches  85;  Indels  14;  Gaps   7;

Db    338 haclsdpchnrgscketslgf-ececspgwtgptcstniddcspnncshggtcq--d-lv 393
          ::|||:|| | |:|:  :: |  | |: |  |  |   |  | :| | |||||:  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    394 ngfkcvcppqwtgktcqldaneceakpcvnaksckcknliasyycdclpgwmgqncdinind 453
          :|| |:|:    |  |::: ::|| : | |  :| : | :| | | |   |: |: :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    454 c---lgqcqndascrdlvngyrcicppgyagdhcerdidecasnpclngghcqneinrfq 510
          |   |  ||:|: |    :|::| | ||| |:||: |:|:|  | | ||:|| : :| :
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    511 clcptgfsgnlcqld---i---dy-cepnpcqnqaqcynrasdyfckcpedyegkncshl 563
          |:|| |:|| :|::   :     |:  |||||||  | :: :| |   |:| :|  |
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155


RESULT   14
ID   W18348 standard; protein; 520 AA.
AC   W18348;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 3; Page 59-61; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   520 AA;

  Query Match           5.7%;  Score 644;  DB 25;  Length 520;
  Best Local Similarity  38.2%;  Pred. No. 2.03e-34;
  Matches   91;  Conservative   53;  Mismatches  81;  Indels  13;  Gaps   9;

Db    275 kpckngatctntgqgsytcscrpgytgatcelgidecdpspcknggsc--td-lensysc 331
          :|||| :|| :     | |:|  |: |  |:: |  | ::|||:||:|   :  |::: |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    332 tcppgfygkicelsamtcadgpcfnggrcsdspdgyscrcpvgysgfncekkidycs-s 390
          |: || |  ||::   | |  | | : | |: :  |:| ||  |:|  || |:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    391 -spcsngakcvdlgdaylcrcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcp 449
          :|| : :||:    :: | | :|: | ||| : |||  : | ||: | |:|| ::| ||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    450 pgytgrnc--sap-v-sr---cehapchngatcherghryvcecargyggpncqfllp 500
          ||:|  |  |:| | :|   |::  |:||| |  | :  :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT   15
ID   W18349 standard; protein; 702 AA.
AC   W18349;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 4; Page 61-64; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   702 AA;

  Query Match           5.7%;  Score 644;  DB 25;  Length 702;
  Best Local Similarity  38.2%;  Pred. No. 2.03e-34;
  Matches   91;  Conservative   53;  Mismatches  81;  Indels  13;  Gaps   9;

Db    275 kpckngatctntgqgsytcscrpgytgatcelgidecdpspcknggsc--td-lensysc 331
          :|||| :|| :     | |:|  |: |  |:: |  | ::|||:||:|   :  |::: |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    332 tcppgfygkicelsamtcadgpcfnggrcsdspdgyscrcpvgysgfncekkidycs-s 390
          |: || |  ||::   | |  | | : | |: :  |:| ||  |:|  || |:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    391 -spcsngakcvdlgdaylcrcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcp 449
          :|| : :||:    :: | | :|: | ||| : |||  : | ||: | |:|| ::| ||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    450 pgytgrnc--sap-v-sr---cehapchngatcherghryvcecargyggpncqfllp 500
          ||:|  |  |:| | :|   |::  |:||| |  | :  :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


Search completed: Fri May 28 08:25:27 1999
Job time : 167 secs.
```

SCANPS (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:25:45 1999; MasPar time 61.26 Seconds
997.388 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-2
Description: (1-1525) from US09191647.pep
Perfect Score: 11299
Sequence: 1 MRGVGWQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 55.224; Variance 119.421; scale 0.462

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4216	37.3	1469	2	B36665	slit protein 2 precur	0.00e+00
2	4210	37.3	1480	2	A36665	slit protein 1 precur	0.00e+00
3	1470	13.0	530	2	A31640	epidermal growth fact	6.88e-246
4	735	6.5	1064	2	A40136	fibropellin Ia - sea	1.20e-105
5	734	6.5	2321	2	S78549	notch3 protein - huma	1.85e-105
6	726	6.4	570	2	A48836	fibropellin C precurs	5.71e-104
7	719	6.4	2318	2	S45306	notch 3 protein - mou	1.15e-102
8	714	6.3	2524	2	A35844	Xotch protein - Afric	9.77e-102
9	708	6.3	2531	2	S18188	notch protein homolog	1.27e-100
10	708	6.3	2703	2	A24420	notch protein - fruit	1.27e-100
11	705	6.2	2471	2	A49128	cell-fate determining	4.60e-100
12	704	6.2	2531	2	A46019	gene Notch-1 protein	7.05e-100
13	666	5.9	728	2	I50719	C-Delta-1 - chicken	7.77e-93
14	667	5.9	2437	2	S42612	transmembrane protein	5.08e-93
15	667	5.9	2555	2	A40043	notch protein homolog	5.08e-93
16	650	5.8	1220	2	A56136	jagged protein precur	7.01e-90
17	647	5.7	1203	2	A49175	Motch B protein - mou	2.51e-89
18	627	5.5	722	2	I48324	DELTA-like 1 - mouse	1.21e-85
19	609	5.4	2139	2	A35672	crumbs protein - frui	2.44e-82
20	554	4.9	832	2	A31246	neurogenic protein De	2.69e-72
21	554	4.9	833	2	S19087	gene Delta protein pr	2.69e-72
22	552	4.9	861	2	A48825	Notch homolog Motch p	6.20e-72
23	554	4.9	880	2	S00670	gene Delta protein pr	2.69e-72
24	558	4.9	1429	2	S06434	homeotic protein lin-	5.04e-73
25	522	4.6	293	2	B26637	neurogenic repetitive	1.68e-66
26	507	4.5	473	2	A56175	adhesive plaque prote	8.47e-64
27	494	4.4	560	2	A60164	platelet membrane gly	1.83e-61
28	486	4.3	1091	2	A58532	glial cell membrane g	4.98e-60
29	470	4.2	387	2	B49175	Motch A protein - mou	3.59e-57
30	478	4.2	1404	2	A36666	serrate protein precu	1.34e-58
31	478	4.2	1408	2	S16148	gene serrate protein	1.34e-58
32	464	4.1	200	2	A26637	neurogenic repetitive	4.21e-56
33	452	4.0	1295	2	A32901	glp1 protein precurso	5.71e-54
34	428	3.8	259	2	S48713	fetal antigen 1 - hum	9.96e-50
35	419	3.7	260	2	A44549	fetal antigen 1 homeo	3.81e-48
36	422	3.7	383	2	B45484	delta-like dlk homeot	1.13e-48
37	422	3.7	383	2	S53716	homeotic protein dlk	1.13e-48
38	422	3.7	1134	1	A29944	chaoptin precursor -	1.13e-48
39	417	3.7	1959	1	AGRT	agrin - rat	8.54e-48
40	405	3.6	385	2	S53718	homeotic protein dlk	1.08e-45
41	404	3.6	4391	2	A38096	perlecan precursor -	1.61e-45
42	391	3.5	360	2	I47020	decorin - rabbit	2.96e-43
43	395	3.5	385	2	A54785	preadipocyte factor 1	5.97e-44
44	395	3.5	536	2	A34901	lysine carboxypeptida	5.97e-44
45	393	3.5	907	2	JE0176	orphan G protein-coup	1.33e-43

ALIGNMENTS

```
RESULT    1
ENTRY           B36665     #type complete
TITLE           slit protein 2 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  16-Dec-1998
ACCESSIONS      B36665
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    B36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1469 ##label ROT
    ##cross-references GB:X53959
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
FEATURE
  66-91           #domain proteoglycan amino-terminal homology #label
                    PAH1\
  101-124         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR1\
  125-148         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR2\
  149-172         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR3\
  173-196         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR4\
  197-220         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR5\
  228-272         #domain proteoglycan carboxyl-terminal homology #label
                    PCS1\
  288-313         #domain proteoglycan amino-terminal homology #label
                    PAH2\
  323-346         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR6\
  347-370         #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LRR7\
  371-394         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494         #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537         #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695         #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733         #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  846-890         #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061       #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match          37.3%;  Score 4216;  DB 2;  Length 1469;
  Best Local Similarity  43.9%;  Pred. No. 0.00e+00;
  Matches  607;  Conservative  328;  Mismatches 394;  Indels  53;  Gaps  34;

Db    105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
          |:|  |::: |    || |  |  |:|  |::: :    |     | |||:|:| | ::
Qy     84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Db    165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
          |: |:||  :::|||| |||:|::: ||::| :||:|||||||:| |:   |  : :||:
Qy    144 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 203

Db    225 LRLSDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLKGQNVADLHDQEFKCSGLTE- 283
          :||  | : |||||:|||  ||   |:: ||:| :||:|:|:|||::: :|| ||   |
Qy    204 FRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEEG 263

Db    284 HAP-MECG-AENSCPHPCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKS 341
          | : |  : :   || :| |:: ||||| |:|| :|  ||:  |::|||| |  :|| :
Qy    264 HQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGA 323

Db    342 FSSFRRLRRIDLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
          ||:::|||||||||| || :| |:|: ||: | :|||||||| :|| ::| || ||:||||
Qy    324 FSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLL 383

Db    402 NANEISCIRKDAFRDLHSLSLLSLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLR 461
          ||| |:|:| |||:|||:|:|||||||::|::|:||| ::::: |:||| |||||||:|:
Qy    384 NANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLK 443

Db    462 WLADYLHKNPIETSGARCESPKRMHRRRIESLREEKFKCS-WGELRMKLSGECRMDSDCP 520
          ||||||| |||||||||| ||:|:  :||  ::  ||:||   : | ||||:|  |  ||
Qy    444 WLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACP 503

Db    521 AMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKR 580
           :|:|||||||: ::|: ||  ||  |:|| ||:||:  : : |:| :||:| |:::
Qy    504 EKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFSN 563

Db    581 NQLTGIEPNAFEGASHIQE--L---QL-G-ENKI-K--E------I-SNKM-------FL 616
          | :| ||  ||||||: |  |   :|   ::|: |  |      : ||::       |:
Qy    564 NKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRITCVGNDSFI 623

Db    617 GLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNLASNPFNCNCHLAWFAECVRKKSLNGG 676
```

```
          ||  :: |:|||||: | ||:|: |:||::||| :||||||| |||::| :||| :  |
Qy    624 GLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKKRIVTG 683

Db    677 AARCGAPSKVRDVQIKDLPHSEFKCSSENSE-GCLGDGYCPPSCTCTGTVVACSRNQLKE 735
          :|| |  :::: | |::  :| |   | : :|   : || ||| ||| || : ||
Qy    684 NPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRCSNKGLKV 743

Db    736 IPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFANLTKLSTL 795
          :|:||| : :||||::|::  :  | : : : || :|||||:|: ||| :|:|:| | ||
Qy    744 LPKGIPRDVTELYLDGNQFTLVPKE-LSNYKHLTLIDLSNNRISTLSNQSFSNMTQLLTL 802

Db    796 IISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSNPLYCDCG 855
          |:|||:|:|:  ::: ||::||::|||| ||::|||:|:|| :|:|:|:|:|||||||
Qy    803 ILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLYCDCN 862

Db    856 LKWFSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQP 915
          : |:|||:| :| |||||||| | :| |||:|:|||  | |:| |  :||||||:|:  |
Qy    863 MQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNP 922

Db    916 CQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLE--EGRFSCQC 973
          | |:: | : |   |:| |  |: |  |:  | || :|||:: :|| : |  |  | | |
Qy    923 CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCIC 982

Db    974 APGYTGARCETNIDDCLGEIKCQNNATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFN 1033
          | |: |  || |:|||  :  |:||:||:||:::| | | | ::||:|: |::||: ::|
Qy    983 ADGFEGENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLN 1041

Db   1034 PCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDD 1093
          || : :||:     : ||| :|: | :|  ::||||:::| ||: | |::| | | ||:
Qy   1042 PCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEG 1101

Db   1094 YTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTS 1153
          |:| :||  :  :|: |:||||:| :| :|   |   : :: :|:| ||| |  || | |
Qy   1102 YSGLFCE-FSP-PMVLPRTSPCDNFDCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVS 1157

Db   1154 ISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELFNGRIRVSYDV 1213
          ::|::::|::::  : : ||::|:|: ::: |::|||:| |:  |:||||: ||:| |||
Qy   1158 VNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASYDT 1217

Db   1214 GNHPVSTMYSFEMVADGKYHAVELLAIKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFL 1273
          |:|| |::|| | : ||::| |||||:  :::| || | :: | | :    |:: :|:::
Qy   1218 GSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYV 1277

Db   1274 GGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQKITPGCAL-LEGE- 1331
          ||:|  :   : :    :| ||| ||::: ||  | || :   |  | |||
Qy   1278 GGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVC 1337

Db   1332 QQEE-EDDEQD-FM-D--ET--PHIKEEPV-DPCLENKCRRGSRCVPNSNARDGYQCKCK 1383
           :  :  | | : |     : ::  |||| ||| :|: |:|  ||  :| |||    .
Qy   1338 AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCL 1394

Db   1384 HGQRGRYCDQAASTCRKEQVREYYTENDCR-SRQPLKYAKCVGGCGNQCCAAKIVRR-RK 1441
          |: |  ||:        |          || |     | | :|   : |   |  |  :
Qy   1395 EGHGGVLCDEEEDLFNPCQAIKC-KHGKCRLSGLGQPYCECSSGYTGDSCDREISCRGER 1453

Db   1442 VR 1443
          :|
Qy   1454 IR 1455
```

```
RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
```

midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession A36665
##status preliminary
##molecule_type mRNA
##residues 1-1480 ##label ROT
##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
KEYWORDS alternative splicing
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR7\
371-394 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR8\
395-418 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR9\
419-442 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR10\
450-494 #domain proteoglycan carboxyl-terminal homology #label PCS2\
512-537 #domain proteoglycan amino-terminal homology #label PAH3\
547-571 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR11\
572-595 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR12\
596-619 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR13\
620-643 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR14\
651-695 #domain proteoglycan carboxyl-terminal homology #label PCS3\
708-733 #domain proteoglycan amino-terminal homology #label PAH4\
743-766 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR15\
767-790 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR16\
791-814 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR17\
815-838 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LR18\
846-890 #domain proteoglycan carboxyl-terminal homology #label PCS4\
1028-1061 #domain EGF homology #label EGF
SUMMARY #length 1480 #molecular-weight 165751 #checksum 900

Query Match 37.3%; Score 4210; DB 2; Length 1480;
Best Local Similarity 44.5%; Pred. No. 0.00e+00;

Matches 600; Conservative 328; Mismatches 369; Indels 52; Gaps 33;

```
Db   105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
Qy    84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Db   165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
Qy   144 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 203

Db   225 LRLSDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLKGQNVADLHDQEFKCSGLTE- 283
Qy   204 FRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEEG 263

Db   284 HAP-MECG-AENSCPHPCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKS 341
Qy   264 HQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGA 323

Db   342 FSSFRRLRRIDLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
Qy   324 FSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLL 383

Db   402 NANEISCIRKDAFRDLHSLSLLSLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLR 461
Qy   384 NANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLK 443

Db   462 WLADYLHKNPIETSGARCESPKRMHRRRIESLREEKFKCS-WGELRMKLSGECRMDSDCP 520
Qy   444 WLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACP 503

Db   521 AMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKR 580
Qy   504 EKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFSN 563

Db   581 NQLTGIEPNAFEGASHIQE--L---QL-G-ENKI-K--E------I-SNKM-------FL 616
Qy   564 NKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRITCVGNDSFI 623

Db   617 GLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNLASNPFNCNCHLAWFAECVRKKSLNGG 676
Qy   624 GLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKKRIVTG 683

Db   677 AARCGAPSKVRDVQIKDLPHSEFKCSSENSE-GCLGDGYCPPSCTCTGTVVACSRNQLKE 735
Qy   684 NPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRCSNKGLKV 743

Db   736 IPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFANLTKLSTL 795
Qy   744 LPKGIPRDVTELYLDGNQFTLVPKE-LSNYKHLTLIDLSNNRISTLSNQSFSNMTQLLTL 802

Db   796 IISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSNPLYCDCG 855
Qy   803 ILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLYCDCN 862

Db   856 LKWFSDWIKLDYVEPGIARCAEPEQMKDKLILSLTPSSSFVCRGRVRNDILAKCNACFEQP 915
Qy   863 MQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNP 922

Db   916 CQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLE--EGRFSCQC 973
Qy   923 CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCIC 982

Db   974 APGYTGARCETNIDDCLGEIKCQNNATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFN 1033
Qy   983 ADGFEGENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLN 1041

Db  1034 PCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDD 1093
Qy  1042 PCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEG 1101

Db  1094 YTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTS 1153
Qy  1102 YSGLFCE-FSP-PMVLPRTSPCDNFDCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVS 1157
```

```
Db   1154 ISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELFNGRIRVSYDV 1213
          ::|::::|:::: : : ||::|:|: ::: |::|||:| |:  |:||||: ||:| |||
Qy   1158 VNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASYDT 1217

Db   1214 GNHPVSTMYSFEMVADGKYHAVELLAIKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFL 1273
          |:|| |::|| | : ||::| |||||:  :::| || | :: | | :    |:: :|:::
Qy   1218 GSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYV 1277

Db   1274 GGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQKITPGCAL-LEGE- 1331
          ||:|  :   : :    :| ||| ||:::: ||  | || :   |  | |||
Qy   1278 GGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVC 1337

Db   1332 QQEE-EDDEQD-FM-D--ET--PHIKEEPV-DPCLENKCRRGSRCVPNSNARDGYQCKCK 1383
           :   :   |  | : |     : ::   |||| ||| :|: |:|  ||  :| |||
Qy   1338 AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCL 1394

Db   1384 HGQRGRYCDQGEGSTEP-PTVT-AASTCR 1410
           |: |  ||: |   :|  ::    : ||
Qy   1395 EGHGGVLCDEEEDLFNPCQAIKCKHGKCR 1423


ULT   3
ENTRY           A31640     #type fragment
TITLE           epidermal growth factor-like protein slit - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A31640
REFERENCE       A31640
  #authors      Rothberg, J.M.; Hartley, D.A.; Walther, Z.;
                  Artavanis-Tsakonas, S.
  #journal      Cell (1988) 55:1047-1059
  #title        slit: An EGF-homologous locus of D. melanogaster involved in
                  the development of the embryonic central nervous system.
  #cross-references MUID:89077533
  #accession    A31640
    ##molecule_type DNA
    ##residues      1-530 ##label ROT
    ##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
  #introns      470/3
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        growth factor
FEATURE
  148-181           #domain EGF homology #label EGF
MARY            #length 530  #checksum 6330

 Query Match           13.0%;  Score 1470;  DB 2;  Length 530;
 Best Local Similarity  39.1%;  Pred. No. 6.88e-246;
 Matches  206;  Conservative  132;  Mismatches 168;  Indels  21;  Gaps  17;

Db      1 MKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHG 60
          | |||:|:|||  | |:| |  :|||||:|:  || |:: | : |   |:| |  |: |
Qy    888 MADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKG 947

Db     61 KHCEFMIDACYGNPCRNNATCTVLE--EGRFSCQCAPGYTGARCETNIDDCLGEIKCQNN 118
           |:  | || :|||:: :|| : |  |  | | || |: |  || |:|||  :  |:||
Qy    948 QDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDC-EDNDCENN 1006

Db    119 ATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFNPCANGAKCMDHFTHYSCDCQAGFHG 178
          :||:||:::| | | | ::||:|: |::||: ::||| : :||:     : ||| :|: |
Qy   1007 STCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVG 1066

Db    179 TNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNH 238
           :|  ::||||:::| ||: | |::| | | ||: |:| :||  :  :|: |:||||:|
Qy   1067 EHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEGYSGLFCE-FSP-PMVLPRTSPCDNF 1124

Db    239 ECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTRPEANVTI 298
          :| :|   |   : :: :|:| ||| |  || | |::|::::|:::: : : ||::|:|:
Qy   1125 DCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITL 1182

Db    299 VFSSGQN-GILMYDGQDAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVELL 357
           ::: :: |||:| |:  |:|||: ||:| ||| |:|| |::|| | : ||::| ||||
Qy   1183 QIATDEDSGILLYKGDKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1242

Db    358 AIKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFLGGLPVDPAQQAYKNWQIRNLTSFKG 417
          |: :::| || | :: | | :    |:: :|:::||:|  :   : :    :| ||| |
Qy   1243 ALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHG 1302

Db    418 CMKEVWINHKLVDFGNAQRQQKITPGCAL-LEGE-QQEE-EDDEQD-FM-D--ET--PHI 468
          |:::: ||  | || :   |  | |||         :   :   |  |  :  |     :
Qy   1303 CIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPL 1362

Db    469 KEEPV-DPCLENKCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQGE 514
           ::   |||| ||| :|: |:|  ||  :| |||  |: |  ||: |
Qy   1363 CDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCLEGHGGVLCDEEE 1406


RESULT   4
ENTRY           A40136     #type complete
TITLE           fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
ALTERNATE_NAMES epidermal growth factor homolog precursor
CONTAINS        alternatively spliced fibropellin Ib (EGFI)
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            13-May-1992 #sequence_revision 17-Sep-1997 #text_change
                  07-Aug-1998
ACCESSIONS      A40136; B40136; C40136; A29316; A43131
REFERENCE       A40136
  #authors      Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff,
                  R.A.
  #journal      J. Mol. Evol. (1989) 29:314-327
  #title        Structural analysis of the uEGF gene in the sea urchin
                  Strongylocentrotus purpuratus reveals more similarity to
                  vertebrate than to invertebrate genes with EGF-like
                  repeats.
  #cross-references MUID:90112459
  #accession    A40136
    ##status       preliminary
    ##molecule_type mRNA
    ##residues      1-114 ##label DEL
    ##cross-references GB:X17530; NID:g10225; PID:g667061
  #accession    B40136
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type DNA
    ##residues      181-251,329-370,'R',372-408,'RA',411-441 ##label DE2
  #accession    C40136
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type DNA
    ##residues      'K',747-821,898-978 ##label DE3
REFERENCE       A29316
  #authors      Hursh, D.A.; Andrews, M.E.; Raff, R.A.
  #journal      Science (1987) 237:1487-1490
  #title        A sea urchin gene encodes a polypeptide homologous to
                  epidermal growth factor.
  #cross-references MUID:87319677
  #accession    A29316
    ##status       preliminary
    ##molecule_type mRNA
    ##residues      'S',280-481,786-1064 ##label HUR
    ##cross-references GB:M17421; NID:g161474; PID:g552260
REFERENCE       A43131
  #authors      Hunt, L.T.; Barker, W.C.
  #journal      FASEB J. (1989) 3:1760-1764
  #title        Avidin-like domain in an epidermal growth factor homolog from
                  a sea urchin.
  #cross-references MUID:89196806
  #contents     annotation
COMMENT         EGF homology repeats 10-17 are spliced out in the short form
                  (fibropellin Ib).
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
```

```
    1-19            #domain signal sequence #status predicted #label SIG\
    20-1064         #product fibropellin I #status predicted #label FIB\
    23-54           #domain EGF homology #label EG01\
    57-175          #domain C1r/C1s repeat homology #label CSR\
    180-211         #domain EGF homology #label EG02\
    218-249         #domain EGF homology #label EG03\
    256-287         #domain EGF homology #label EG04\
    294-325         #domain EGF homology #label EG05\
    332-363         #domain EGF homology #label EG06\
    370-401         #domain EGF homology #label EG07\
    408-439         #domain EGF homology #label EG08\
    446-477         #domain EGF homology #label EG09\
    484-515         #domain EGF homology #label EG10\
    522-553         #domain EGF homology #label EG11\
    560-591         #domain EGF homology #label EG12\
    598-629         #domain EGF homology #label EG13\
    636-667         #domain EGF homology #label EG14\
    674-705         #domain EGF homology #label EG15\
    712-743         #domain EGF homology #label EG16\
    750-781         #domain EGF homology #label EG17\
    788-819         #domain EGF homology #label EG18\
    826-857         #domain EGF homology #label EG19\
    864-895         #domain EGF homology #label EG20\
    902-933         #domain EGF homology #label EG21\
    936-1064        #region avidin-like\
    23-34,28-43,45-54,
    62-88,180-191,
    185-200,202-211,
    218-229,223-238,
    240-249,256-267,
    261-276,278-287,
    294-305,299-314,
    316-325,332-343,
    337-352,354-363,
    370-381,375-390,
    392-401,408-419,
    413-428,430-439,
    446-457,451-466,
    468-477,484-495    #disulfide_bonds #status predicted\
    489-504,506-515,
    522-533,527-542,
    544-553,560-571,
    565-580,582-591,
    598-609,603-618,
    620-629,636-647,
    641-656,658-667,
    674-685,679-694,
    696-705,712-723,
    717-732,734-743,
    750-761,755-770,
    772-781,788-799,
    793-808,810-819,
    826-837,831-846,
    848-857,864-875,
    869-884,886-895,
    902-913,907-922,
    924-933         #disulfide_bonds #status predicted
SUMMARY         #length 1064  #molecular-weight 112072  #checksum 303

  Query Match           6.5%;  Score 735;  DB 2;  Length 1064;
  Best Local Similarity  44.5%;  Pred. No. 1.20e-105;
  Matches  106;  Conservative  38;  Mismatches  81;  Indels  13;  Gaps  9;

Db    254 NECASSPCLNGGIC-VDGVNMFECTCLAGFTGVRCEVNIDECASAPCQNGGIC-I-DGI- 309
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    310 NGYTCSCPLGFSGDNCENNDDECSSIPCLNGGTCVDLVNAYMCVCAPGWTGPTCADNIDE 369
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    370 CA-S-APCQNGGVCIDGVNGYMCDCQPGYTGTHCETDIDECARPPCQNGGDCVDGVNGYV 427
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    428 CICAPGFDGLNCE-NN--I----DECASRPCQNGAVCVDGVNGFVCTCSAGYTGVLCE 478
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT    5
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
   #authors     Joutel, A.; Tournier-Lasserve, E.
   #submission  submitted to the EMBL Data Library, April 1997
   #accession   S78549
      ##molecule_type mRNA
      ##residues      1-2321 ##label JOU1
      ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE       S71825
   #authors     Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
   #journal     Nature (1996) 383:707-710
   #title       Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
   #cross-references MUID:97032728
   #accession   S71825
      ##status       nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues     67-113;138-194;268-333,'G',335-346;536-613;716-765;
                      1240-1279;1815-1888 ##label JOU2
      ##cross-references EMBL:U97669
GENETICS
   #gene        notch3
   #map_position 19p13.1
FUNCTION
   #description  may be involved in pathogenesis of CADASIL, causing a type of
                   stroke and dementia
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        tandem repeat; transmembrane protein
FEATURE
   318-349          #domain EGF homology #label EGF\
   1838-1870        #domain ankyrin repeat homology #label AN1\
   1871-1903        #domain ankyrin repeat homology #label AN2\
   1905-1937        #domain ankyrin repeat homology #label AN3\
   1938-1970        #domain ankyrin repeat homology #label AN4\
   1971-2003        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2321  #molecular-weight 243657  #checksum 3337

  Query Match           6.5%;  Score 734;  DB 2;  Length 2321;
  Best Local Similarity  31.5%;  Pred. No. 1.85e-105;
  Matches  170;  Conservative  100;  Mismatches 227;  Indels  42;  Gaps  36;

Db    160 DECRVGEPCRHGGTCLNTPGSF-RCQCPAGYTGPLCENPAVPCAPSPCRNGGTCR-QSGD 217
Qy    916 NPC-LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    218 LT-YDCACLPGFEGQNCEVNVDDCPGHRCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVD 276
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    277 ECQLQPNACHNGGTCFNTLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASF 336
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    337 YCACPMGKTGLLCHLDDACV---SNPCHEDAICDTNP---VN-GRAICTCPPGFTGGACD 389
```

```
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153

Db    390 QDVDECSIGANPCEHLGRC-VNTQGSFLCQCGRGY-TGPR-CETDVNECLSGPCRNQA-T 445

Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRA 1213

Db    446 CLDRIGQFTC-IC-MAGFT-GTYCEVDI-DECQS-SPCVNGGVCKDRVN-GFSCT-C-PS 497

Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDS 1273

Db    498 G-FSGSTC-QLDVDECASTPCRNGAKCVD-QPDGYECRCAEGFEGTLCDRNV-DDCSP-D 552

Qy   1274 PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCH 1333

Db    553 P--CHHGRC-VDGIASFSCACAPGYTGTRCESQVDE-CRSQPCRHGGKCLDLVD-KYLCR 607

Qy   1334 KKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 1392

Db    608 CPSGTTGVNCEVNID--D-CASNPCTFGVCR-DGINR-YDCVCQPGFTGPLCNVEINEC 661

Qy   1393 CLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449


RESULT    6
ENTRY            A48836     #type complete
TITLE            fibropellin C precursor - sea urchin (Strongylocentrotus
                   purpuratus)
ALTERNATE_NAMES  EGF repeat-containing protein; epidermal growth
                   factor-related protein 3; fibropellin III
ORGANISM         #formal_name Strongylocentrotus purpuratus #common_name
                   purple urchin
DATE             01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                   07-Aug-1998
ACCESSIONS       A48836
REFERENCE        A48836
   #authors      Bisgrove, B.W.; Raff, R.A.
   #journal      Dev. Biol. (1993) 157:526-538
   #title        The SpEGF III gene encodes a member of the fibropellins: EGF
                   repeat-containing proteins that form the apical lamina of
                   the sea urchin embryo.
   #cross-references MUID:93273088
   #accession    A48836
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-570 ##label BIS
      ##cross-references GB:L07045; NID:g310659; PID:g310660
      ##note         sequence extracted from NCBI backbone (NCBIN:132724,
                      NCBIP:132725)
CLASSIFICATION   #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-18             #domain signal sequence #status predicted #label SIG\
   19-570           #product fibropellin C #status predicted #label FIB\
   19-54            #domain EGF homology #label EGF1\
   57-175           #domain C1r/C1s repeat homology #label C1R\
   176-211          #domain EGF homology #label EGF2\
   214-249          #domain EGF homology #label EGF3\
   252-287          #domain EGF homology #label EGF4\
   290-325          #domain EGF homology #label EGF5\
   328-363          #domain EGF homology #label EGF6\
   366-401          #domain EGF homology #label EGF7\
   404-439          #domain EGF homology #label EGF8\
   442-570          #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439    #disulfide_bonds #status predicted
SUMMARY          #length 570  #molecular-weight 61115  #checksum 5567

  Query Match           6.4%;  Score 726;  DB 2;  Length 570;
  Best Local Similarity  40.3%;  Pred. No. 5.71e-104;
  Matches   96;  Conservative   54;  Mismatches  75;  Indels  13;  Gaps   8;

Db    178 DDCTPNPCLNGATC-VDQVNDYQCICAPGFTGDNCETDIDECASAPCRNGGAC-V-D-QV 233

Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    234 NGYTCNCIPGFNGVNCENNINECASIPCLNGGICVDGINQFACTCLPGYTGILCETDINE 293

Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    294 CA-S-SPCQNGGSCTDAVNRYTCDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYV 351

Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    352 CQCLPNYTGTHCEIS----L---DACASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402

Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT    7
ENTRY            S45306     #type complete
TITLE            notch 3 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   10-Jul-1998
ACCESSIONS       S45306
REFERENCE        S45306
   #authors      Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal      Mech. Dev. (1994) 46:123-136
   #title        The novel Notch homologue mouse Notch 3 lacks specific
                   epidermal growth factor-repeats and is expressed in
                   proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession    S45306
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1839-1871        #domain ankyrin repeat homology #label AN1\
   1872-1904        #domain ankyrin repeat homology #label AN2\
   1906-1938        #domain ankyrin repeat homology #label AN3\
   1939-1971        #domain ankyrin repeat homology #label AN4\
   1972-2004        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           6.4%;  Score 719;  DB 2;  Length 2318;
  Best Local Similarity  31.2%;  Pred. No. 1.15e-102;
  Matches  168;  Conservative  100;  Mismatches 229;  Indels  42;  Gaps  35;

Db    161 DECRSGTTCRHGGTCLNTPGSF-RCQCPLGYTGLLCENPVVPCAPSPCRNGGTC--RQSS 217

Qy    916 NPCLSNP-CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    218 DVTYDCACLPGFEGQNCEVNVDDCPGHRCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVD 277

Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    278 ECQLQPNACHNGGTCFNLLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASF 337

Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    338 YCACPMGKTGLLCHLDDACV---SNPCHEDAICDTNP---VS-GRAICTCPPGFTGGACD 390

Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153
```

```
Db    391 QDVDECSIGANPCEHLGRC-VNTQGSFLCQCGRGY-TGPR-CETDVNECLSGPCRNQA-T 446
Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRA 1213

Db    447 CLDRIGQFTC-IC-MAGFT-GTYCEVDI-DECQS-SPCVNGGVCKDRVN-GFSCT-C-PS 498
Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDS 1273

Db    499 G-FSGSMC-QLDVDECASTPCRNGAKCVD-QPDGYECRCAEGFEGTLCERNV-DDCSP-D 553
Qy   1274 PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCH 1333

Db    554 P--CHHGRC-VDGIASFSCACAPGYTGIRCESQVDE-CRSQPCRYGGKCLDLVD-KYLCR 608
Qy   1334 KKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 1392

Db    609 CPPGTTGVNCEVNID--D-CASNPCTFGVCR-DGINR-YDCVCQPGFTGPLCNVEINEC 662
Qy   1393 CLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449


RESULT    8
ENTRY           A35844     #type complete
TITLE           Xotch protein - African clawed frog
ORGANISM        #formal_name Xenopus laevis #common_name African clawed frog
DATE            12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A35844
REFERENCE       A35844
   #authors     Coffman, C.; Harris, W.; Kintner, C.
   #journal     Science (1990) 249:1438-1441
   #title       Xotch, the Xenopus homolog of Drosophila notch.
   #cross-references MUID:90385285
   #accession   A35844
      ##status       preliminary; nucleic acid sequence not shown; not
                      compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-2524 ##label COF
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        transmembrane protein
FEATURE
   222-254          #domain EGF homology #label EGF\
   1924-1956        #domain ankyrin repeat homology #label AN1\
   1957-1989        #domain ankyrin repeat homology #label AN2\
   1991-2023        #domain ankyrin repeat homology #label AN3\
   2024-2056        #domain ankyrin repeat homology #label AN4\
   2057-2089        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2524  #molecular-weight 274931  #checksum 9441

  Query Match           6.3%;  Score 714;  DB 2;  Length 2524;
  Best Local Similarity  31.1%;  Pred. No. 9.77e-102;
  Matches  170;  Conservative  105;  Mismatches 227;  Indels  45;  Gaps  34;

Db    182 NECSQNPCKNGGQCINE-FGSYRCTCQNRFTGRNCDEPYVPCNPSPCLNGGTC--RQTDD 238
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    239 TSYDCTCLPGFSGQNCEENIDDCPSNNCRNGGTCVDGVNTYNCQCPPDWTGQYCTEDVDE 298
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    299 CQLMPNACQNGGTCHNTYGGYNCVCVNGWTGEDCSENIDDCANAACHSGATCHDRVASFY 358
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    359 CECPHGRTGLLCHLDNA-CI--SNPCNEGSNCDTNP---VN-GKAICTCPPGYTGPACNN 411
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCEK 1154

Db    412 DVDECSLGANPC-E-RGGRC-TNTLGSFQCNCPQ--G---YAGPRCEIDVNECLSNPCQN 463
Qy   1155 LVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRAS 1214

Db    464 -DSTCLDQIGEFQCICMPGYEGLYCET-NIDECASNPC-LHNGKCIDKIN-E--FRCDCP 517
Qy   1215 YDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSP 1274

Db    518 TGFSGNLC-QHDFDECTSTPCKNGAKCLDG--PNSYTCQCTEGFTGRHCEQDI-NECIP- 572
Qy   1275 L-YVGGMPGKSNVASLRQAPGQNGTS-FHGCIRNLYINSELQDFQKVPMQTGILPGCEPC 1332

Db    573 DP--CHYGTCK-DGIATFTCLCRPGYTGRLCDNDINE-CLSKPCLNGGQCTDREN-GYIC 627
Qy   1333 HKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPINAFSYSC 1391

Db    628 TCPKGTTGVNCETKID--D-CASNLCDNGKC-IDKI-DGYECTCEPGYTGKLCNININEC 682
Qy   1392 KCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449

Db    683 DSNPCRN 689
Qy   1450 RGERIRD 1456


RESULT    9
ENTRY           S18188     #type complete
TITLE           notch protein homolog - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
                  12-Feb-1999
ACCESSIONS      S18188
REFERENCE       S18188
   #authors     Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal     Development (1991) 113:199-205
   #title       A homolog of Drosophila Notch expressed during mammalian
                  development.
   #cross-references MUID:92111383
   #accession   S18188
      ##molecule_type mRNA
      ##residues     1-2531 ##label WEI
      ##cross-references EMBL:X57405; NID:g57634; PID:g57635
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1917-1949        #domain ankyrin repeat homology #label AN1\
   1950-1982        #domain ankyrin repeat homology #label AN2\
   1984-2016        #domain ankyrin repeat homology #label AN3\
   2017-2049        #domain ankyrin repeat homology #label AN4\
   2050-2082        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2531  #molecular-weight 270907  #checksum 2705

  Query Match           6.3%;  Score 708;  DB 2;  Length 2531;
  Best Local Similarity  29.1%;  Pred. No. 1.27e-100;
  Matches  161;  Conservative  116;  Mismatches 229;  Indels  48;  Gaps  38;

Db    420 ANPCEHAGKCL-NTLGSFECQCLQGYTGPRCEIDVNECISNPCQNDATC-L-D-QIGEFQ 475
Qy    920 SNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFW 979

Db    476 CICMPGYEGVYCEINTDECASSPCLHNGRCVDKINEFLCQCPKGFSGHLCQYDVDECA-S 534
Qy    980 CICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQD 1039

Db    535 -TPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHIGL-CKDGVATFTCLCQ 592
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    593 PGYTGHHCE-TN--I----NECHSQPCRHGGTCQDRDNYYLCLCLKGTTGPNCEINLDDC 645
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE-KL--V 1156

Db    646 ASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPCHNGGTCEDGIAGFTCRCP 705
Qy   1157 SVNFINKESYLQ-IPSAKVRPQTNIT--L-QIATDEDSGILLYKGDKDHIAVELYRGRVR 1212
```

```
Db    706 EGYHDPTCLSEVNECNSNPCIHGACRD-GLNGYKCDCAPGWSGTNCDINNNEC-ESNPCV 763
          :| |    :     : :    |  :   | :     : :  | |  |  |   :|
Qy   1213 ASY-DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNF 1271

Db    764 NGGTCKD-MTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTCIDDVAGYKCNCPLPY 821
          ::      | : : | : |:: :| |  |::: |  |  :| :  ::|   :  : |:
Qy   1272 DSPLYVGGMPGKSNVASLRQA-PGQN-GTSFHGCIRNLYIN-SE-LQD---FQ-KVPMQ- 1322

Db    822 TGATCEVVLAPCATSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEIDINE-CVKSPCRH 880
          ||        ||    | : | |  | :  :|:| |  || |  |:   |: |: : | |
Qy   1323 TGILPGC--EPCHKKVCAH-GTCQPSSQ-AGFTCECQEGWMGPLCDQRTNDPCLGNKCVH 1378

Db    881 GASCQNTNG-SYRCLCQAGYTGRNCESDID--D-CRPNPCHNGGSCT-DGVNAAFCDCLP 935
          |: |   |: || | |  |  |  |: : |  : |::  | :|  |   |:  ::|:| :
Qy   1379 GT-CLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRLSGLGQPYCECSS 1436

Db    936 GFQGAFCEEDINEC 949
          |: |  |: :|: |
Qy   1437 GYTGDSCDREIS-C 1449


RESULT  10
ENTRY           A24420     #type complete
TITLE           notch protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES neurogenic repetitive locus protein
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
                  07-Aug-1998
ACCESSIONS      A24420; A24768; S09358; A05267
REFERENCE       A24420
  #authors      Kidd, S.; Kelley, M.R.; Young, M.W.
  #journal      Mol. Cell. Biol. (1986) 6:3094-3108
  #cross-references MUID:87064624
  #accession    A24420
    ##molecule_type DNA
    ##residues      1-2703 ##label KID
    ##cross-references GB:K03508; NID:g157991; PID:g157993
REFERENCE       A24768
  #authors      Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
  #journal      Cell (1985) 43:567-581
  #cross-references MUID:86079539
  #accession    A24768
    ##molecule_type mRNA
    ##residues      1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',
                    268-872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',
                    2258-2264,'V',2266-2406,'R',2408-2444,'L',2446-2703
                    ##label WHA1
    ##note          the authors translated the codon ATC for residue 49 as
                    Thr, ATT for residue 2044 as Arg, GTA for residue 2265
                    as Ala, CGC for residue 2407 as His, and CTT for
                    residue 2445 as Arg
REFERENCE       S09358
  #authors      Tautz, D.
  #journal      Nucleic Acids Res. (1989) 17:6463-6471
  #title        Hypervariability of simple sequences as a general source for
                  polymorphic DNA markers.
  #cross-references MUID:89385974
  #accession    S09358
    ##molecule_type DNA
    ##residues      2505-2551,'QQQQ',2552-2576,'E',2578-2604 ##label TAU
REFERENCE       A05267
  #authors      Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.;
                  Artavanis-Tsakonas, S.
  #journal      Cell (1985) 40:55-62
  #title        opa: a novel family of transcribed repeats shared by the
                  Notch locus and other developmentally regulated loci in D.
                  melanogaster.
  #cross-references MUID:85099329
  #accession    A05267
    ##molecule_type DNA
    ##residues      2504-2576,'E',2578-2611 ##label WHA2
GENETICS
  #gene         notch; opa
    ##cross-references FlyBase:FBgn0004647
  #map_position 8.96-9.36
  #introns      53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
CLASSIFICATION  #superfamily notch protein; ankyrin repeat homology; EGF
                  homology
KEYWORDS        differentiation; tandem repeat; transmembrane protein
FEATURE
  27-43           #domain transmembrane #status predicted #label TMM1\
  568-599         #domain EGF homology #label EGF\
  1746-1762       #domain transmembrane #status predicted #label TMM2\
  1950-1982       #domain ankyrin repeat homology #label AN1\
  1983-2015       #domain ankyrin repeat homology #label AN2\
  1988-2004       #domain transmembrane #status predicted #label TMM3\
  2017-2049       #domain ankyrin repeat homology #label AN3\
  2050-2082       #domain ankyrin repeat homology #label AN4\
  2083-2115       #domain ankyrin repeat homology #label AN5\
  2538-2568       #region glutamine-rich\
  2538-2568       #domain neurogenic repetitive element #status predicted
                    #label OPA
SUMMARY         #length 2703  #molecular-weight 288876  #checksum 6404

  Query Match           6.3%;  Score 708;  DB 2;  Length 2703;
  Best Local Similarity  44.8%;  Pred. No. 1.27e-100;
  Matches  108;  Conservative   37;  Mismatches  82;  Indels  14;  Gaps   7;

Db    490 NECESHPCQNEGSCLDDPGTF-RCVCMPGFTGTQCEIDIDECQSNPCLNDGTCH--D-KI 545
          | | |:|| |:|:|  ||  | || |  || | :|:: |  | |||| : ||||  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    546 NGFKCSCALGFTGARCQINIDDCQSQPCRNRGICHDSIAGYSCECPPGYTGTSCEININD 605
          :|| | || || |  |::|:|||:   | | : | |:|  |:| ||| |||  || :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    606 C--DSNPCHR-GKCIDDVNSFKCLCDPGYTGYICQKQINECESNPCQFDGHCQDRVGSYY 662
          |  | |||:: :|||   ::||| | ||| |  |: ::::|: | |   :|| | | :|
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    663 CQCQAGTSGKNCE-----V--NVNECHSNPCNNGATCIDGINSYKCQCVPGFTGQHCEKN 715
          | |  | ||  ||     |    : | :  | ||| ||  ||   |||:||: |: |||
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155

Db    716 V 716
          |
Qy   1156 V 1156


RESULT  11
ENTRY           A49128     #type complete
TITLE           cell-fate determining gene Notch2 protein - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A49128
REFERENCE       A49128
  #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
  #journal      Development (1992) 116:931-941
  #title        Notch2: a second mammalian Notch gene.
  #cross-references MUID:93202015
  #accession    A49128
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type mRNA
    ##residues      1-2471 ##label WEI
    ##experimental_source Schwann cell
    ##note         sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
  1029-1060       #domain EGF homology #label EGF\
  1876-1908       #domain ankyrin repeat homology #label AN1\
  1909-1941       #domain ankyrin repeat homology #label AN2\
  1943-1975       #domain ankyrin repeat homology #label AN3\
  1976-2008       #domain ankyrin repeat homology #label AN4\
  2009-2041       #domain ankyrin repeat homology #label AN5
```

```
SUMMARY         #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           6.2%;  Score 705;  DB 2;  Length 2471;
  Best Local Similarity  29.5%;  Pred. No. 4.60e-100;
  Matches   162;  Conservative  118;  Mismatches 220;  Indels  49;  Gaps  38;

Db    184 NECDIPGRCQHGGTCLNLPGS-YRCQCPQRFTGQHCDSPYVPCAPSPCVNGGTCR-QTGD 241
Qy    916 NPC-LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    242 FTSE-CHCLPGFEGSNCERNIDDCPNHKCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVD 300
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    301 ECLLQPNACQNGGTCTNRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASF 360
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    361 SCLCPEGKAGLLCH------L--DDACISNPCHKGALCDTNPLNGQYICTCPQAYKGADC 412
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQC-IVRINEP-ICQCLPGYQGEKC 1152

Db    413 TEDVDECAMANSNPC-E-HAGKC-VNTDGAFHCEC-LK-G---YAGPRCEMDINECHSDP 464
Qy   1153 -EKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRV 1211

Db    465 CQN-DATCLDKIGGFTCLCMP-G-FKGVHCELEVNECQSNPC-VNNGQCVDKVNRFQCLC 520
Qy   1212 RASYDTGSHPASAIYSVETINDGNFHIVE-LLALDQSLSLSVDGGNPKIITNLSKQSTLN 1270

Db    521 --PPGFTGPVC-QIDIDDCSSTPCLNGAKCIDH-PNGY-ECQCATGF-TGTLCDENIDNC 574
Qy   1271 FDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSE-LQDFQKVPMQTGILPGC 1329

Db    575 DP-DP--CHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 628
Qy   1330 EPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 1388

Db    629 YQCNCQPGTSGLNC--EIN-FDDCASNPCLHGAC-VDGINR-YSCVCSPGFTGQRCNIDI 683
Qy   1389 YSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREI 1447

Db    684 DECASNPCR 692
Qy   1448 S-CRGERIR 1455


RESULT   12
ENTRY           A46019     #type complete
TITLE           gene Notch-1 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A46019
REFERENCE       A46019
   #authors     del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins,
                  N.A.; Copeland, N.G.; Gridley, T.
   #journal     Genomics (1993) 15:259-264
   #title       Cloning, analysis, and chromosomal localization of Notch-1, a
                  mouse homolog of Drosophila Notch.
   #cross-references MUID:93194170
   #accession   A46019
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues     1-2531 ##label DEL
      ##cross-references GB:Z11886; GB:S47228; NID:g288502; PID:g288503
      ##note         sequence extracted from NCBI backbone (NCBIP:127318)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   757-788          #domain EGF homology #label EGF\
   1917-1948        #domain ankyrin repeat homology #label AN1\
   1949-1981        #domain ankyrin repeat homology #label AN2\
   1983-2015        #domain ankyrin repeat homology #label AN3\
   2016-2048        #domain ankyrin repeat homology #label AN4\
   2049-2081        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2531  #molecular-weight 271312  #checksum 6611

  Query Match           6.2%;  Score 704;  DB 2;  Length 2531;
  Best Local Similarity  28.5%;  Pred. No. 7.05e-100;
  Matches   158;  Conservative  118;  Mismatches 230;  Indels  48;  Gaps  38;

Db    420 ANRCEHAGKCL-NTLGSFECQCLQGYTGPGCEIDVNECISNPCQNDATC-L-D-QIGEFQ 475
Qy    920 SNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFW 979

Db    476 CICMPGYEGVYCEINTDECASSPCLHNGHCMDKIHEFQCQCPKGFNGHLCQYDVDECA-S 534
Qy    980 CICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQD 1039

Db    535 -TPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGS-CKDGVATFTCLCQ 592
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    593 PGYTGHHCE-TN--I----NECHSQPCRHGGTCQDRDNSYLCLCLKGTTGPNCEINLDDC 645
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE-KL--V 1156

Db    646 ASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPCHNGGTCEDGIAGFTCRCP 705
Qy   1157 SVNFINKESYLQ-IPSAKVRPQTNIT--L-QIATDEDSGILLYKGDKDHIAVELYRGRVR 1212

Db    706 EGYHDPTCLSEVNECNSNPCIHGACRD-GLNGYKCDCAPGWSGTNCDINNNEC-ESNPCV 763
Qy   1213 ASY-DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNF 1271

Db    764 NGGTCKD-MTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTCIDDVAGYKCNCPLPY 821
Qy   1272 DSPLYVGGMPGKSNVASLRQA-PGQN-GTSFHGCIRNLYIN-SE-LQD---FQ-KVPMQ- 1322

Db    822 TGATCEVVLAPCATSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEVDINE-CVKSPCRH 880
Qy   1323 TGILPGC--EPCHKKVCAH-GTCQPSSQ-AGFTCECQEGWMGPLCDQRTNDPCLGNKCVH 1378

Db    881 GASCQNTNG-SYRCLCQAGYTGRNCESDID--D-CRPNPCHNGGSCT-DGINTAFCDCLP 935
Qy   1379 GT-CLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRLSGLGQPYCECSS 1436

Db    936 GFQGAFCEEDINEC 949
Qy   1437 GYTGDSCDREIS-C 1449


RESULT   13
ENTRY           I50719     #type complete
TITLE           C-Delta-1 - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  14-Aug-1998
ACCESSIONS      I50719
REFERENCE       I50719
   #authors     Henrique, D.; Adam, J.; Myat, A.; Chitnis, A.; Lewis, J.;
                  Ish-Horowicz, D.
   #journal     Nature (1995) 375:787-790
   #title       Expression of a Delta homologue in prospective neurons in the
                  chick.
   #cross-references MUID:95319507
   #accession   I50719
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-728 ##label HEN
      ##cross-references EMBL:U26590; NID:g882411; PID:g882412
CLASSIFICATION  #superfamily EGF homology
FEATURE
   454-485          #domain EGF homology #label EGF
SUMMARY         #length 728  #molecular-weight 79861  #checksum 1765
```

```
  Query Match           5.9%;  Score 666;  DB 2;  Length 728;
  Best Local Similarity  39.5%;  Pred. No. 7.77e-93;
  Matches   94;  Conservative   50;  Mismatches  81;  Indels  13;  Gaps   9;

Db    303 KPCKNGATCTNTGQGSYTCSCRPGYTGSSCEIEINECDANPCKNGGSC--TD-LENSYSC 359
          :|||| :|| :     | |:|  |: |  |:: |: | :||||:||:|   :  |::: |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    360 TCPPGFYGKNCELSAMTCADGPCFNGGRCTDNPDGGYSCRCPLGYSGFNCEKKIDYCS-S 418
          |: || | |||::   | |  | | : | |  :  |:| ||  |:|  || |:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    419 -SPCANGAQCVDLGNSYICQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCP 477
           :|| : : |:   ::: |:| :|: | ||| : |||    | ||: | |:|| |:| ||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    478 PGYNGKNC--STP-V-SR---CEHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
          ||:|  |  | | | :|   |::  |:||| |  | |  :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT   14
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
  #authors      Bierkamp, C.; Campos-Ortega, J.A.
  #journal      Mech. Dev. (1993) 43:87-100
  #title        A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
  #cross-references MUID:94128602
  #accession    S42612
    ##status        preliminary
    ##molecule_type mRNA
    ##residues      1-2437 ##label BIE
    ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
  1915-1947          #domain ankyrin repeat homology #label AN1\
  1948-1980          #domain ankyrin repeat homology #label AN2\
  1982-2014          #domain ankyrin repeat homology #label AN3\
  2015-2047          #domain ankyrin repeat homology #label AN4\
  2048-2080          #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           5.9%;  Score 667;  DB 2;  Length 2437;
  Best Local Similarity  39.3%;  Pred. No. 5.08e-93;
  Matches   94;  Conservative   53;  Mismatches  79;  Indels  13;  Gaps   9;

Db    791 NECASNPCLNQGSCIDDVAGF-KCNCMLPYTGEVCENVLAPCSPRPCKNGGVC--RESED 847
          | | |||| |:|:|  |   | :| |   : |: |:  : :| : |||:|| |  :|:|:
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    848 FQSFSCNCPAGWQGQTCEVDINECVRNPCTNGGVCENLRGGFQCRCNPGFTGALCENDID 907
          ::| | |: | :|: |||::::|  | | | : | :    : | | | :|| |||: :|
Qy    976 -DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    908 DC--EPNPCSNGGVCQDRVNGFVCVCLAGFRGERCAEDIDECVSAPCRNGGNCTDCVNSY 965
          |  : ||| : : |    :|| | | :|: ||:|  |:|:|    |:||::||| ||:|
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    966 TCSCPAGFSGINCEINTPDC-TESS-C--FN--GGT-CVDGISSFSCVCLPGFTGNYCQ 1017
          || || |:||: ||:: |     :| |  |:   |: |:  |:   | ||||: |: |:
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT   15
ENTRY           A40043     #type complete
TITLE           notch protein homolog TAN-1 precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A40043
REFERENCE       A40043
  #authors      Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds,
                  T.C.; Smith, S.D.; Sklar, J.
  #journal      Cell (1991) 66:649-661
  #title        TAN-1, the human homolog of the Drosophila Notch gene, is
                  broken by chromosomal translocations in T lymphoblastic
                  neoplasms.
  #cross-references MUID:91347367
  #accession    A40043
    ##status        preliminary; nucleic acid sequence not shown; not
                     compared with conceptual translation
    ##molecule_type mRNA
    ##residues      1-2555 ##label ELL
    ##cross-references GB:M73980
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
  1149-1180          #domain EGF homology #label EGF\
  1927-1959          #domain ankyrin repeat homology #label AN1\
  1960-1992          #domain ankyrin repeat homology #label AN2\
  1994-2026          #domain ankyrin repeat homology #label AN3\
  2027-2059          #domain ankyrin repeat homology #label AN4\
  2060-2092          #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2555  #molecular-weight 272337  #checksum 463

  Query Match           5.9%;  Score 667;  DB 2;  Length 2555;
  Best Local Similarity  30.7%;  Pred. No. 5.08e-93;
  Matches  169;  Conservative  106;  Mismatches 225;  Indels  51;  Gaps  45;

Db    302 MPNACQNGGTCKNTHGG-YNCVCVNGWTGEDCSENIDDCASAACFHGATCH--D-RVASF 357
          ::|:| | |||::    | | |  |  |:||   |  | | :| ||:|||  :    :|
Qy    919 LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGF 978

Db    358 YCECPHGRTGLLCHLN-DACISNPCNEGSNCDTNPVNGKAICTCPSGYTGPACSQDVDEC 416
          | |: |  |  | :| | |  | |:: | |  : :| :  | ||: |||  | : :| |
Qy    979 WCICADGFEGENCEVNVDDCEDNDCENNSTC-VDGIN-NYTCLCPPEYTGELCEEKLDFC 1036

Db    417 SLGANPCEHAGKCINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQC 476
          :  |||:| :||| |  :| |:|  || | :|:|| ::|  | | | | | :  : |
Qy   1037 AQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTC 1096

Db    477 MCMPGYEGVHCEVNTDEC-A-SSPC--L--HNG-RCLDKINEFQCECPTGFTGHLCQDVD 529
          :|  || |: || :     : :|||  :  :|| :|: :|||  |:|  |: |  |: :
Qy   1097 ICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLV 1156

Db    530 ECASTPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCL 589
                |: :  |: |::   |  :    |  ::  || |:    | : || :|    |
Qy.  1157 SVNFIN-KE-S-YLQIPSA-K-VRPQTNI-TL-QIATDE-DSGILLYKGDKDHIAV-E-L 1206

Db    590 CRPGYTGHHCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGTTGPNCEINLDDCASSPC 649
           | |      :|  :  :|    :  |  |  | :    |  :   :  :::|  :::|
Qy   1207 YR-GRVRASYDTGSHP-ASAIYSV-ETIND-GN-FHIVELL-ALDQSLSLSVD--GGNP- 1257

Db    650 DSGTCLDKIDGYECACEPGYTGSMC-NSNIDECAGNPCHNGGTCEDGINGFTCRCP-EGY 707
            | | |    :   | | |:|  :||:      | :|| :   |  :      : :
Qy   1258 KIITNLSKQSTLNFDS-PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDF 1316

Db    708 HD-P--T-CLSEVNECNSNPCVHGACR-DSLNGYKCDCDPGWSGTNCDINNNE-CESNPC 761
          :  |  |  |:  : |: : | ||:|:  |  |: |:|: || |  ||´ |: | :| |
Qy   1317 QKVPMQTGILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKC 1376

Db    762 VNGGTCKDMTS-GIVCTCREGFSGPNC--QTNI-NECASNPCLNKGTC-IDDVA-GYKCN 815
          |:| ||  : : :  | | || :|  |  : :: | | :  | : | | :  ::  | |:
Qy   1377 VHG-TCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKC-KHGKCRLSGLGQPY-CE 1433

Db    816 CLLPYTGATCE 826
          |  ||| :|:
```

Qy 1434 CSSGYTGDSCD 1444

Search completed: Fri May 28 08:28:42 1999
Job time : 177 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:29:00 1999; MasPar time 42.51 Seconds
1014.108 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-2
Description: (1-1525) from US09191647.pep
Perfect Score: 11299
Sequence: 1 MRGVGWQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 56.708; Variance 101.314; scale 0.560

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4210	37.3	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	0.00e+00
2	735	6.5	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	1.25e-127
3	726	6.4	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	1.32e-125
4	719	6.4	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	4.93e-124
5	714	6.3	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	6.53e-123
6	708	6.3	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	1.45e-121
7	708	6.3	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	1.45e-121
8	704	6.2	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	1.14e-120
9	667	5.9	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	2.16e-112
10	667	5.9	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	2.16e-112
11	646	5.7	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	1.03e-107
12	631	5.6	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	2.22e-104
13	627	5.5	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	1.71e-103
14	622	5.5	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	2.20e-102
15	609	5.4	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	1.67e-99
16	554	4.9	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	2.15e-87
17	558	4.9	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	2.85e-88
18	494	4.4	560	1	GPV_HUMAN	PLATELET GLYCOPROTEIN	2.47e-74
19	490	4.3	567	1	GPV_MOUSE	PLATELET GLYCOPROTEIN	1.80e-73
20	478	4.2	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	6.98e-71
21	456	4.0	567	1	GPV_RAT	PLATELET GLYCOPROTEIN	3.67e-66
22	452	4.0	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	2.63e-65
23	422	3.7	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	6.31e-59
24	422	3.7	1134	1	CHAO_DROME	CHAOPTIN PRECURSOR (PH	6.31e-59
25	417	3.7	1959	1	AGRI_RAT	AGRIN PRECURSOR.	7.20e-58
26	405	3.6	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	2.43e-55
27	395	3.5	536	1	CBP8_HUMAN	CARBOXYPEPTIDASE N 83	3.06e-53
28	393	3.5	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	8.02e-53
29	370	3.3	357	1	PGS2_CHICK	BONE PROTEOGLYCAN II P	4.94e-48
30	375	3.3	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	4.54e-49
31	373	3.3	360	1	PGS2_CANFA	BONE PROTEOGLYCAN II P	1.18e-48
32	377	3.3	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	1.74e-49
33	366	3.2	359	1	PGS2_HUMAN	BONE PROTEOGLYCAN II P	3.33e-47
34	358	3.2	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	1.49e-45
35	365	3.2	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	5.36e-47
36	351	3.1	354	1	PGS2_MOUSE	BONE PROTEOGLYCAN II P	4.10e-44
37	346	3.1	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	4.34e-43
38	351	3.1	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	4.10e-44
39	342	3.0	354	1	PGS2_RAT	BONE PROTEOGLYCAN II P	2.86e-42
40	340	3.0	361	1	CHAD_BOVIN	CHONDROADHERIN PRECURS	7.32e-42
41	335	3.0	368	1	PGS1_HUMAN	BONE/CARTILAGE PROTEOG	7.64e-41
42	336	3.0	369	1	PGS1_RAT	BONE/CARTILAGE PROTEOG	4.78e-41
43	341	3.0	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	4.57e-42
44	333	2.9	369	1	PGS1_MOUSE	BONE/CARTILAGE PROTEOG	1.95e-40
45	333	2.9	682	1	CONN_DROME	CONNECTIN PRECURSOR.	1.95e-40

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE  1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          37.3%;  Score 4210;  DB 1;  Length 1480;
  Best Local Similarity  44.5%;  Pred. No. 0.00e+00;
  Matches  600;  Conservative  328;  Mismatches 369;  Indels  52;  Gaps  33;

Db   105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
Qy    84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Db   165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
Qy   144 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 203

Db   225 LRLSDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLKGQNVADLHDQEFKCSGLTE- 283
Qy   204 FRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEEG 263

Db   284 HAP-MECG-AENSCPHPCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKS 341
Qy   264 HQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGA 323

Db   342 FSSFRRLRRIDLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
Qy   324 FSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLL 383

Db   402 NANEISCIRKDAFRDLHSLSLLSLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLR 461
Qy   384 NANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLK 443

Db   462 WLADYLHKNPIETSGARCESPKRMHRRRIESLREEKFKCS-WGELRMKLSGECRMDSDCP 520
Qy   444 WLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACP 503

Db   521 AMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKR 580
Qy   504 EKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFSN 563

Db   581 NQLTGIEPNAFEGASHIQE--L---QL-G-ENKI-K--E------I-SNKM-------FL 616
Qy   564 NKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRITCVGNDSFI 623

Db   617 GLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNLASNPFNCNCHLAWFAECVRKKSLNGG 676
Qy   624 GLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKKRIVTG 683

Db   677 AARCGAPSKVRDVQIKDLPHSEFKCSSENSE-GCLGDGYCPPSCTCTGTVVACSRNQLKE 735
Qy   684 NPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRCSNKGLKV 743

Db   736 IPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFANLTKLSTL 795
```

```
Qy    744 LPKGIPRDVTELYLDGNQFTLVPKE-LSNYKHLTLIDLSNNRISTLSNQSFSNMTQLLTL 802

Db    796 IISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSNPLYCDCG 855

Qy    803 ILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLYCDCN 862

Db    856 LKWFSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQP 915

Qy    863 MQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNP 922

Db    916 CQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLE--EGRFSCQC 973

Qy    923 CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCIC 982

Db    974 APGYTGARCETNIDDCLGEIKCQNNATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFN 1033

Qy    983 ADGFEGENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLN 1041

Db   1034 PCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDD 1093

Qy   1042 PCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEG 1101

Db   1094 YTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTS 1153

Qy   1102 YSGLFCE-FSP-PMVLPRTSPCDNFDCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVS 1157

Db   1154 ISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELFNGRIRVSYDV 1213

Qy   1158 VNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASYDT 1217

Db   1214 GNHPVSTMYSFEMVADGKYHAVELLAIKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFL 1273

Qy   1218 GSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYV 1277

Db   1274 GGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQKITPGCAL-LEGE- 1331

Qy   1278 GGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVC 1337

Db   1332 QQEE-EDDEQD-FM-D--ET--PHIKEEPV-DPCLENKCRRGSRCVPNSNARDGYQCKCK 1383

Qy   1338 AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCL 1394

Db   1384 HGQRGRYCDQGEGSTEP-PTVT-AASTCR 1410

Qy   1395 EGHGGVLCDEEEDLFNPCQAIKCKHGKCR 1423
```

```
RESULT    2
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1064       FIBROPELLIN I.
FT   DOMAIN      20     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480    516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      518    554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      556    592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      594    630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      632    668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      670    706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      708    744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      746    782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      784    820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      822    858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      860    896       EGF-LIKE 20.
FT   DOMAIN      898    934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      936   1064       AVIDIN-LIKE.
FT   DISULFID     23     34       BY SIMILARITY.
FT   DISULFID     28     43       BY SIMILARITY.
FT   DISULFID     45     54       BY SIMILARITY.
FT   DISULFID    180    191       BY SIMILARITY.
FT   DISULFID    185    200       BY SIMILARITY.
FT   DISULFID    202    211       BY SIMILARITY.
FT   DISULFID    218    229       BY SIMILARITY.
FT   DISULFID    223    238       BY SIMILARITY.
FT   DISULFID    240    249       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    261    276       BY SIMILARITY.
FT   DISULFID    278    287       BY SIMILARITY.
FT   DISULFID    294    305       BY SIMILARITY.
FT   DISULFID    299    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   DISULFID    446    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    477       BY SIMILARITY.
FT   DISULFID    484    495       BY SIMILARITY.
FT   DISULFID    489    504       BY SIMILARITY.
FT   DISULFID    506    515       BY SIMILARITY.
FT   DISULFID    522    533       BY SIMILARITY.
FT   DISULFID    527    542       BY SIMILARITY.
FT   DISULFID    544    553       BY SIMILARITY.
FT   DISULFID    560    571       BY SIMILARITY.
FT   DISULFID    565    580       BY SIMILARITY.
FT   DISULFID    582    591       BY SIMILARITY.
FT   DISULFID    598    609       BY SIMILARITY.
FT   DISULFID    603    618       BY SIMILARITY.
FT   DISULFID    620    629       BY SIMILARITY.
FT   DISULFID    636    647       BY SIMILARITY.
FT   DISULFID    641    656       BY SIMILARITY.
FT   DISULFID    658    667       BY SIMILARITY.
FT   DISULFID    674    685       BY SIMILARITY.
FT   DISULFID    679    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    712    723       BY SIMILARITY.
FT   DISULFID    717    732       BY SIMILARITY.
FT   DISULFID    734    743       BY SIMILARITY.
FT   DISULFID    750    761       BY SIMILARITY.
FT   DISULFID    755    770       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    788    799       BY SIMILARITY.
FT   DISULFID    793    808       BY SIMILARITY.
FT   DISULFID    810    819       BY SIMILARITY.
FT   DISULFID    826    837       BY SIMILARITY.
FT   DISULFID    831    846       BY SIMILARITY.
FT   DISULFID    848    857       BY SIMILARITY.
FT   DISULFID    864    875       BY SIMILARITY.
FT   DISULFID    869    884       BY SIMILARITY.
FT   DISULFID    886    895       BY SIMILARITY.
FT   DISULFID    902    913       BY SIMILARITY.
FT   DISULFID    907    922       BY SIMILARITY.
FT   DISULFID    924    933       BY SIMILARITY.
FT   VARSPLIC    477    780       MISSING (IN FORM IB).
FT   CARBOHYD     30     30       POTENTIAL.
FT   CARBOHYD    136    136       POTENTIAL.
FT   CARBOHYD    851    851       POTENTIAL.
FT   CONFLICT    279    279       L -> S (IN REF. 2).
SQ   SEQUENCE   1064 AA;  112072 MW;  FBD10D48 CRC32;

  Query Match           6.5%;  Score 735;  DB 1;  Length 1064;
  Best Local Similarity  44.5%;  Pred. No. 1.25e-127;
  Matches  106;  Conservative   38;  Mismatches  81;  Indels  13;  Gaps   9;

Db    254 NECASSPCLNGGIC-VDGVNMFECTCLAGFTGVRCEVNIDECASAPCQNGGIC-I-DGI- 309
          | | |:|| | | |  | |: : |||  || |  |:| | | | || :|| | : :|
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    310 NGYTCSCPLGFSGDNCENNDDECSSIPCLNGGTCVDLVNAYMCVCAPGWTGPTCADNIDE 369
          :|: | |: || |:||| | |:|    | | :|||| :| | |:|:|  ||  | :::|
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    370 CA-S-APCQNGGVCIDGVNGYMCDCQPGYTGTHCETDIDECARPPCQNGGDCVDGVNGYV 427
          ||  · |||: : ||   :|::||| ||| | ||: |:|:|    | ||: | |:||||
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    428 CICAPGFDGLNCE-NN--I----DECASRPCQNGAVCVDGVNGFVCTCSAGYTGVLCE 478
          |||: |: || || :   :      | :  ||||| |:  :|  :| | :|| |  ||
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT    3
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
```

```
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   570       FIBROPELLIN C.
FT   DOMAIN      18    55       EGF-LIKE 1.
FT   DOMAIN      62   175       CUB.
FT   DOMAIN     176   212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214   250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252   288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290   326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328   364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366   402       EGF-LIKE 7.
FT   DOMAIN     404   440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442   570       AVIDIN-LIKE.
FT   DISULFID    23    34       BY SIMILARITY.
FT   DISULFID    28    43       BY SIMILARITY.
FT   DISULFID    45    54       BY SIMILARITY.
FT   DISULFID   180   191       BY SIMILARITY.
FT   DISULFID   185   200       BY SIMILARITY.
FT   DISULFID   202   211       BY SIMILARITY.
FT   DISULFID   218   229       BY SIMILARITY.
FT   DISULFID   223   238       BY SIMILARITY.
FT   DISULFID   240   249       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   261   276       BY SIMILARITY.
FT   DISULFID   278   287       BY SIMILARITY.
FT   DISULFID   294   305       BY SIMILARITY.
FT   DISULFID   299   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   CARBOHYD    30    30       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match           6.4%;  Score 726;  DB 1;  Length 570;
  Best Local Similarity  40.3%;  Pred. No. 1.32e-125;
  Matches   96;  Conservative   54;  Mismatches  75;  Indels  13;  Gaps   8;

Db    178 DDCTPNPCLNGATC-VDQVNDYQCICAPGFTGDNCETDIDECASAPCRNGGAC-V-D-QV 233
          : | :||| | :||  | |: |:| |: || |::|:  |  | | ||::||:| : : :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    234 NGYTCNCIPGFNGVNCENNINECASIPCLNGGICVDGINQFACTCLPGYTGILCETDINE 293
          :|: | |  ||:| ||| |:::|    | | : ||||| ::| | | ||| |||  ::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    294 CA-S-SPCQNGGSCTDAVNRYTCDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYV 351
          ||   :|||: : |  : : : ||| :|: | :|: ::::|  : | ||: | |:|:||
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    352 CQCLPNYTGTHCEIS----L---DACASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402
          | |   |:|  ||:|    |    :| :: ||||: |     : :|:||||| | :||
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153
```

```
RESULT    4
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643       EXTRACELLULAR.
FT   TRANSMEM  1644   1664       POTENTIAL.
FT   DOMAIN    1665   2318       CYTOPLASMIC.
FT   DOMAIN      39   1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261       PEST.
FT   DOMAIN      39     78       EGF-LIKE 1.
FT   DOMAIN      79    119       EGF-LIKE 2.
FT   DOMAIN     120    157       EGF-LIKE 3.
FT   DOMAIN     159    196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235       EGF-LIKE 5.
FT   DOMAIN     237    273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313       EGF-LIKE 7.
FT   DOMAIN     315    351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390       EGF-LIKE 9.
FT   DOMAIN     392    430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      696    731       EGF-LIKE 18.
FT   DOMAIN      735    771       EGF-LIKE 19.
FT   DOMAIN      772    809       EGF-LIKE 20.
FT   DOMAIN      811    848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      850    886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      888    923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      925    961       EGF-LIKE 24.
FT   DOMAIN      963    999       EGF-LIKE 25.
FT   DOMAIN     1001   1035       EGF-LIKE 26.
FT   DOMAIN     1037   1083       EGF-LIKE 27.
FT   DOMAIN     1085   1121       EGF-LIKE 28.
FT   DOMAIN     1123   1159       EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1161   1204       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1206   1245       EGF-LIKE 31.
FT   DOMAIN     1247   1288       EGF-LIKE 32.
FT   DOMAIN     1290   1326       EGF-LIKE 33.
FT   DOMAIN     1336   1374       EGF-LIKE 34.
FT   REPEAT     1388   1428       LIN/NOTCH 1.
FT   REPEAT     1429   1467       LIN/NOTCH 2.
FT   REPEAT     1468   1503       LIN/NOTCH 3.
FT   REPEAT     1784   1816       CDC10/SWI6 1.
FT   REPEAT     1817   1865       CDC10/SWI6 2.
FT   REPEAT     1866   1898       CDC10/SWI6 3.
FT   REPEAT     1899   1932       CDC10/SWI6 4.
FT   REPEAT     1933   1965       CDC10/SWI6 5.
FT   REPEAT     1966   1998       CDC10/SWI6 6.
FT   DISULFID     43     55       BY SIMILARITY.
FT   DISULFID     49     66       BY SIMILARITY.
FT   DISULFID     68     77       BY SIMILARITY.
FT   DISULFID     83     94       BY SIMILARITY.
FT   DISULFID     88    107       BY SIMILARITY.
FT   DISULFID    109    118       BY SIMILARITY.
FT   DISULFID    124    135       BY SIMILARITY.
FT   DISULFID    129    145       BY SIMILARITY.
FT   DISULFID    147    156       BY SIMILARITY.
FT   DISULFID    163    175       BY SIMILARITY.
FT   DISULFID    169    184       BY SIMILARITY.
FT   DISULFID    186    195       BY SIMILARITY.
FT   DISULFID    202    213       BY SIMILARITY.
FT   DISULFID    207    223       BY SIMILARITY.
FT   DISULFID    225    234       BY SIMILARITY.
FT   DISULFID    241    252       BY SIMILARITY.
FT   DISULFID    246    261       BY SIMILARITY.
FT   DISULFID    263    272       BY SIMILARITY.
FT   DISULFID    279    292       BY SIMILARITY.
FT   DISULFID    286    301       BY SIMILARITY.
FT   DISULFID    303    312       BY SIMILARITY.
FT   DISULFID    319    330       BY SIMILARITY.
FT   DISULFID    324    339       BY SIMILARITY.
FT   DISULFID    341    350       BY SIMILARITY.
FT   DISULFID    356    367       BY SIMILARITY.
FT   DISULFID    361    378       BY SIMILARITY.
FT   DISULFID    380    389       BY SIMILARITY.
FT   DISULFID    396    409       BY SIMILARITY.
FT   DISULFID    403    418       BY SIMILARITY.
FT   DISULFID    420    429       BY SIMILARITY.
FT   DISULFID    436    447       BY SIMILARITY.
FT   DISULFID    441    456       BY SIMILARITY.
FT   DISULFID    458    467       BY SIMILARITY.
FT   DISULFID    474    485       BY SIMILARITY.
FT   DISULFID    479    494       BY SIMILARITY.
FT   DISULFID    496    505       BY SIMILARITY.
FT   DISULFID    512    523       BY SIMILARITY.
FT   DISULFID    517    532       BY SIMILARITY.
FT   DISULFID    534    543       BY SIMILARITY.
FT   DISULFID    550    560       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    571    580       BY SIMILARITY.
FT   DISULFID    587    598       BY SIMILARITY.
FT   DISULFID    592    607       BY SIMILARITY.
FT   DISULFID    609    618       BY SIMILARITY.
FT   DISULFID    625    635       BY SIMILARITY.
FT   DISULFID    630    644       BY SIMILARITY.
FT   DISULFID    646    655       BY SIMILARITY.
FT   DISULFID    662    673       BY SIMILARITY.
FT   DISULFID    667    682       BY SIMILARITY.
FT   DISULFID    684    693       BY SIMILARITY.
FT   DISULFID    700    710       BY SIMILARITY.
FT   DISULFID    705    719       BY SIMILARITY.
FT   DISULFID    721    730       BY SIMILARITY.
FT   DISULFID    739    750       BY SIMILARITY.
FT   DISULFID    744    759       BY SIMILARITY.
FT   DISULFID    761    770       BY SIMILARITY.
FT   DISULFID    776    787       BY SIMILARITY.
FT   DISULFID    781    797       BY SIMILARITY.
FT   DISULFID    799    808       BY SIMILARITY.
FT   DISULFID    815    827       BY SIMILARITY.
FT   DISULFID    821    836       BY SIMILARITY.
FT   DISULFID    838    847       BY SIMILARITY.
FT   DISULFID    854    865       BY SIMILARITY.
FT   DISULFID    859    874       BY SIMILARITY.
FT   DISULFID    876    885       BY SIMILARITY.
FT   DISULFID    892    902       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    913    922       BY SIMILARITY.
FT   DISULFID    929    940       BY SIMILARITY.
FT   DISULFID    934    949       BY SIMILARITY.
FT   DISULFID    951    960       BY SIMILARITY.
FT   DISULFID    967    978       BY SIMILARITY.
FT   DISULFID    972    987       BY SIMILARITY.
FT   DISULFID    989    998       BY SIMILARITY.
FT   DISULFID   1005   1016       BY SIMILARITY.
FT   DISULFID   1010   1023       BY SIMILARITY.
FT   DISULFID   1025   1034       BY SIMILARITY.
FT   DISULFID   1041   1062       BY SIMILARITY.
FT   DISULFID   1056   1071       BY SIMILARITY.
FT   DISULFID   1073   1082       BY SIMILARITY.
FT   DISULFID   1089   1100       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.
..;
Note: remainder of annotations omitted.

  Query Match           6.4%;  Score 719;  DB 1;  Length 2318;
  Best Local Similarity  31.2%;  Pred. No. 4.93e-124;
  Matches  168;  Conservative  100;  Mismatches 229;  Indels  42;  Gaps  35;

Db    161 DECRSGTTCRHGGTCLNTPGSF-RCQCPLGYTGLLCENPVVPCAPSPCRNGGTC--RQSS 217
          : | |   |:: ||| : |  | || || |: |  |: |: :| ::||::|||| :::
Qy    916 NPCLSNP-CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    218 DVTYDCACLPGFEGQNCEVNVDDCPGHRCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVD 277
          :  : | |  ||||:|||||||||  : | | :|||||:| | | |||| ||::| | :|
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    278 ECQLQPNACHNGGTCFNLLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASF 337
            | : |:|:: : |:   | | |  | || |  ::|||   | :|| | | | ::
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094
```

```
Db    338 YCACPMGKTGLLCHLDDACV---SNPCHEDAICDTNP---VS-GRAICTCPPGFTGGACD 390
          | || | :||:| :  : |   ::|| ::  |: :   |    :|| | ||: | |:
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153

Db    391 QDVDECSIGANPCEHLGRC-VNTQGSFLCQCGRGY-TGPR-CETDVNECLSGPCRNQA-T 446
            |    | ::  ::    | | ::  | :    :|      | :       | :  :
Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRA 1213

Db    447 CLDRIGQFTC-IC-MAGFT-GTYCEVDI-DECQS-SPCVNGGVCKDRVN-GFSCT-C-PS 498
          | :: :  | :  :  | :  |::    || |  |:|| |   | :   |    |
Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDS 1273

Db    499 G-FSGSMC-QLDVDECASTPCRNGAKCVD-QPDGYECRCAEGFEGTLCERNV-DDCSP-D 553
            : |:|    :|     :| :|||:       : |     : |:    :  :   | |
Qy   1274 PLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCH 1333

          554 P--CHHGRC-VDGIASFSCACAPGYTGIRCESQVDE-CRSQPCRYGGKCLDLVD-KYLCR 608
            | || |   : |:|:| | |  |  |  |: : :: | :  |  |  || :    | |:
Qy   1334 KKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 1392

Db    609 CPPGTTGVNCEVNID--D-CASNPCTFGVCR-DGINR-YDCVCQPGFTGPLCNVEINEC 662
          |  |  || |: : |  : | :  |  | ||  |: : | | | :|:||  |: ||: |
Qy   1393 CLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449
```

RESULT 5
ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN XOTCH.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90385285.
RA COFFMAN C., HARRIS W., KINTNER C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL SCIENCE 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
[redacted]
KINTNER C.;
SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
[redacted]
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M33874; G1364263; -.
DR PIR; A35844; A35844.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS01187; EGF_CA; 21.
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR HSSP; P00740; 1IXA.
KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.

FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	2524	NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT	DOMAIN	20	1728	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1729	1750	POTENTIAL.
FT	DOMAIN	1751	2524	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	57	EGF-LIKE 1.
FT	DOMAIN	58	99	EGF-LIKE 2.
FT	DOMAIN	102	140	EGF-LIKE 3.
FT	DOMAIN	141	177	EGF-LIKE 4.
FT	DOMAIN	179	215	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	217	254	EGF-LIKE 6.
FT	DOMAIN	256	292	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	294	332	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	334	370	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	371	409	EGF-LIKE 10.
FT	DOMAIN	411	449	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	451	487	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	489	525	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	527	563	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	565	600	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	602	638	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	640	675	EGF-LIKE 17.
FT	DOMAIN	677	713	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	715	750	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	752	788	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	790	826	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	828	866	EGF-LIKE 22.
FT	DOMAIN	868	904	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	906	942	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	944	980	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	982	1018	EGF-LIKE 26.
FT	DOMAIN	1020	1056	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28.
FT	DOMAIN	1096	1142	EGF-LIKE 29.
FT	DOMAIN	1144	1180	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	DOMAIN	1441	1560	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1441	1478	LIN/NOTCH 1.
FT	REPEAT	1479	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1560	LIN/NOTCH 3.
FT	DOMAIN	1871	2083	6 X ANK MOTIF REPEATS.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.

```
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   375   386       BY SIMILARITY.
FT   DISULFID   380   397       BY SIMILARITY.
FT   DISULFID   399   408       BY SIMILARITY.
FT   DISULFID   415   428       BY SIMILARITY.
FT   DISULFID   422   437       BY SIMILARITY.
FT   DISULFID   439   448       BY SIMILARITY.
FT   DISULFID   455   466       BY SIMILARITY.
FT   DISULFID   460   475       BY SIMILARITY.
FT   DISULFID   477   486       BY SIMILARITY.
FT   DISULFID   493   504       BY SIMILARITY.
FT   DISULFID   498   513       BY SIMILARITY.
FT   DISULFID   515   524       BY SIMILARITY.
FT   DISULFID   531   542       BY SIMILARITY.
FT   DISULFID   536   551       BY SIMILARITY.
FT   DISULFID   553   562       BY SIMILARITY.
FT   DISULFID   569   579       BY SIMILARITY.
FT   DISULFID   574   588       BY SIMILARITY.
FT   DISULFID   590   599       BY SIMILARITY.
FT   DISULFID   606   617       BY SIMILARITY.
FT   DISULFID   611   626       BY SIMILARITY.
FT   DISULFID   628   637       BY SIMILARITY.
FT   DISULFID   644   654       BY SIMILARITY.
FT   DISULFID   649   663       BY SIMILARITY.
FT   DISULFID   665   674       BY SIMILARITY.
FT   DISULFID   681   692       BY SIMILARITY.
FT   DISULFID   686   701       BY SIMILARITY.
FT   DISULFID   703   712       BY SIMILARITY.
FT   DISULFID   719   729       BY SIMILARITY.
FT   DISULFID   724   738       BY SIMILARITY.
FT   DISULFID   740   749       BY SIMILARITY.
FT   DISULFID   756   767       BY SIMILARITY.
FT   DISULFID   761   776       BY SIMILARITY.
FT   DISULFID   778   787       BY SIMILARITY.
FT   DISULFID   794   805       BY SIMILARITY.
FT   DISULFID   799   814       BY SIMILARITY.
FT   DISULFID   816   825       BY SIMILARITY.
FT   DISULFID   832   843       BY SIMILARITY.
FT   DISULFID   837   854       BY SIMILARITY.
FT   DISULFID   856   865       BY SIMILARITY.
FT   DISULFID   872   883       BY SIMILARITY.
FT   DISULFID   877   892       BY SIMILARITY.
FT   DISULFID   894   903       BY SIMILARITY.
FT   DISULFID   910   921       BY SIMILARITY.
FT   DISULFID   915   930       BY SIMILARITY.
FT   DISULFID   932   941       BY SIMILARITY.
FT   DISULFID   986   997       BY SIMILARITY.
FT   DISULFID   991  1006       BY SIMILARITY.
FT   DISULFID  1008  1017       BY SIMILARITY.
FT   DISULFID  1024  1035       BY SIMILARITY.
FT   DISULFID  1029  1044       BY SIMILARITY.
FT   DISULFID  1046  1055       BY SIMILARITY.
FT   DISULFID  1062  1073       BY SIMILARITY.
FT   DISULFID  1067  1082       BY SIMILARITY.
FT   DISULFID  1084  1093       BY SIMILARITY.
FT   DISULFID  1100  1121       BY SIMILARITY.
FT   DISULFID  1115  1130       BY SIMILARITY.
FT   DISULFID  1132  1141       BY SIMILARITY.
FT   DISULFID  1148  1159       BY SIMILARITY.
FT   DISULFID  1153  1168       BY SIMILARITY.
FT   DISULFID  1170  1179       BY SIMILARITY.
FT   DISULFID  1186  1197       BY SIMILARITY.
FT   DISULFID  1191  1206       BY SIMILARITY.
FT   DISULFID  1208  1217       BY SIMILARITY.
FT   DISULFID  1224  1243       BY SIMILARITY.
FT   DISULFID  1237  1252       BY SIMILARITY.
FT   DISULFID  1254  1263       BY SIMILARITY.
FT   DISULFID  1270  1283       BY SIMILARITY.
FT   DISULFID  1275  1292       BY SIMILARITY.
FT   DISULFID  1294  1303       BY SIMILARITY.
FT   DISULFID  1310  1321       BY SIMILARITY.
FT   DISULFID  1315  1333       BY SIMILARITY.
FT   DISULFID  1335  1344       BY SIMILARITY.
FT   DISULFID  1351  1362       BY SIMILARITY.
FT   DISULFID  1356  1371       BY SIMILARITY.
FT   DISULFID  1373  1382       BY SIMILARITY.
FT   DISULFID  1390  1401       BY SIMILARITY.
FT   DISULFID  1395  1412       BY SIMILARITY.
FT   DISULFID  1414  1423       BY SIMILARITY.
FT   CARBOHYD   462   462       POTENTIAL.
FT   CARBOHYD   887   887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match           6.3%;  Score 714;  DB 1;  Length 2524;
  Best Local Similarity  31.1%;  Pred. No. 6.53e-123;
  Matches  170;  Conservative  105;  Mismatches 227;  Indels  45;  Gaps  34;

Db    181 NECSQNPCKNGGQCINE-FGSYRCTCQNRFTGRNCDEPYVPCNPSPCLNGGTC--RQTDD 237
          | |  ||||| | | ::    |||||   | |::|| |  :| ::|| :||||  :: ::
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    238 TSYDCTCLPGFSGQNCEENIDDCPSNNCRNGGTCVDGVNTYNCQCPPDWTGQYCTEDVDE 297
          :: | |  || |:||| |:|||  |:| | :|||||:| | | |||: ||: | | :|
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    298 CQLMPNACQNGGTCHNTYGGYNCVCVNGWTGEDCSENIDDCANAACHSGATCHDRVASFY 357
          |    |:||: : |  |  |::| |  |  || |  ::||| :  | :|| | | | ::
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    358 CECPHGRTGLLCHLDNA-CI--SNPCNEGSNCDTNP---VN-GKAICTCPPGYTGPACNN 410
          | || | :|||:| :  :  :  ::|| :  :|:  :   |    :|| | ||| |  |::
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPC-DNFDCQNGAQCIVRINEPICQCLPGYQGEKCEK 1154

Db    411 DVDECSLGANPC-E-HGGRC-TNTLGSFQCNCPQ--G---YAGPRCEIDVNECLSNPCQN 462
          |    :  ::  :  :::    |  ::|    :  |   | | :  | |:  :    :
Qy   1155 LVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRAS 1214

Db    463 -DSTCLDQIGEFQCICMPGYEGLYCET-NIDECASNPC-LHNGKCIDKIN-E--FRCDCP 516
           |:      : :    :   :    |   :|:  | :    | | | ::: :  :  | |
Qy   1215 YDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSP 1274

Db    517 TGFSGNLC-QHDFDECTSTPCKNGAKCLDG--PNSYTCQCTEGFTGRHCEQDI-NECIP- 571
            : | :   :      :|  ||:  : |   | |     : |     :  |   | |
Qy   1275 L-YVGGMPGKSNVASLRQAPGQNGTS-FHGCIRNLYINSELQDFQKVPMQTGILPGCEPC 1332

Db    572 DP--CHYGTCK-DGIATFTCLCRPGYTGRLCDNDINE-CLSKPCLNGGQCTDREN-GYIC 626
            |  |||   : | ||| |: |  | |||   |: ||:: |::|  |   :  :| |
Qy   1333 HKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGT-CLPINAFSYSC 1391

Db    627 TCPKGTTGVNCETKID--D-CASNLCDNGKC-IDKI-DGYECTCEPGYTGKLCNININEC 681
          |  |  || |:   |  : | :  | :||| :  : : | | | :|||  |: :|: |
Qy   1392 KCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREIS-C 1449

Db    682 DSNPCRN 688
            ::  |:
Qy   1450 RGERIRD 1456


RESULT   6
ID   NTC1_RAT      STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
```

```
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19  2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19  1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724  1746       POTENTIAL.
FT   DOMAIN    1747  2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    58       EGF-LIKE 1.
FT   DOMAIN      59    99       EGF-LIKE 2.
FT   DOMAIN     102   139       EGF-LIKE 3.
FT   DOMAIN     140   176       EGF-LIKE 4.
FT   DOMAIN     178   216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218   255       EGF-LIKE 6.
FT   DOMAIN     257   293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295   333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335   371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372   410       EGF-LIKE 10.
FT   DOMAIN     412   450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452   488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490   526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528   564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566   601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603   639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641   676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678   714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716   751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753   789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791   827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829   867       EGF-LIKE 22.
FT   DOMAIN     869   905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907   943       EGF-LIKE 24.
FT   DOMAIN     945   981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983  1019       EGF-LIKE 26.
FT   DOMAIN    1021  1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059  1095       EGF-LIKE 28.
FT   DOMAIN    1097  1143       EGF-LIKE 29.
FT   DOMAIN    1145  1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183  1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221  1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267  1305       EGF-LIKE 33.
FT   DOMAIN    1307  1346       EGF-LIKE 34.
FT   DOMAIN    1348  1384       EGF-LIKE 35.
FT   DOMAIN    1387  1426       EGF-LIKE 36.
FT   DOMAIN    1449  1462       CYS-RICH.
FT   DOMAIN    1865  2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865  1910       ANK MOTIF 1.
FT   REPEAT    1912  1942       ANK MOTIF 2.
FT   REPEAT    1944  1975       ANK MOTIF 3.
FT   REPEAT    1978  2009       ANK MOTIF 4.
FT   REPEAT    2011  2042       ANK MOTIF 5.
FT   REPEAT    2044  2076       ANK MOTIF 6.
FT   DISULFID    24    37       BY SIMILARITY.
FT   DISULFID    31    46       BY SIMILARITY.
FT   DISULFID    48    57       BY SIMILARITY.
FT   DISULFID    63    74       BY SIMILARITY.
FT   DISULFID    68    87       BY SIMILARITY.
FT   DISULFID    89    98       BY SIMILARITY.
FT   DISULFID   106   117       BY SIMILARITY.
FT   DISULFID   111   127       BY SIMILARITY.
FT   DISULFID   129   138       BY SIMILARITY.
FT   DISULFID   144   155       BY SIMILARITY.
FT   DISULFID   149   164       BY SIMILARITY.
FT   DISULFID   166   175       BY SIMILARITY.
FT   DISULFID   182   195       BY SIMILARITY.
FT   DISULFID   189   204       BY SIMILARITY.
FT   DISULFID   206   215       BY SIMILARITY.
FT   DISULFID   222   233       BY SIMILARITY.
FT   DISULFID   227   243       BY SIMILARITY.
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   261   272       BY SIMILARITY.
FT   DISULFID   266   281       BY SIMILARITY.
FT   DISULFID   283   292       BY SIMILARITY.
FT   DISULFID   299   312       BY SIMILARITY.
FT   DISULFID   306   321       BY SIMILARITY.
FT   DISULFID   323   332       BY SIMILARITY.
FT   DISULFID   339   350       BY SIMILARITY.
FT   DISULFID   344   359       BY SIMILARITY.
FT   DISULFID   361   370       BY SIMILARITY.
FT   DISULFID   376   387       BY SIMILARITY.
FT   DISULFID   381   398       BY SIMILARITY.
FT   DISULFID   400   409       BY SIMILARITY.
FT   DISULFID   416   429       BY SIMILARITY.
FT   DISULFID   423   438       BY SIMILARITY.
FT   DISULFID   440   449       BY SIMILARITY.
FT   DISULFID   456   467       BY SIMILARITY.
FT   DISULFID   461   476       BY SIMILARITY.
FT   DISULFID   478   487       BY SIMILARITY.
FT   DISULFID   494   505       BY SIMILARITY.
FT   DISULFID   499   514       BY SIMILARITY.
FT   DISULFID   516   525       BY SIMILARITY.
FT   DISULFID   532   543       BY SIMILARITY.
FT   DISULFID   537   552       BY SIMILARITY.
FT   DISULFID   554   563       BY SIMILARITY.
FT   DISULFID   570   580       BY SIMILARITY.
FT   DISULFID   575   589       BY SIMILARITY.
FT   DISULFID   591   600       BY SIMILARITY.
FT   DISULFID   607   618       BY SIMILARITY.
FT   DISULFID   612   627       BY SIMILARITY.
FT   DISULFID   629   638       BY SIMILARITY.
FT   DISULFID   645   655       BY SIMILARITY.
FT   DISULFID   650   664       BY SIMILARITY.
FT   DISULFID   666   675       BY SIMILARITY.
FT   DISULFID   682   693       BY SIMILARITY.
FT   DISULFID   687   702       BY SIMILARITY.
FT   DISULFID   704   713       BY SIMILARITY.
FT   DISULFID   720   730       BY SIMILARITY.
FT   DISULFID   725   739       BY SIMILARITY.
FT   DISULFID   741   750       BY SIMILARITY.
FT   DISULFID   757   768       BY SIMILARITY.
FT   DISULFID   762   777       BY SIMILARITY.
FT   DISULFID   779   788       BY SIMILARITY.
FT   DISULFID   795   806       BY SIMILARITY.
FT   DISULFID   800   815       BY SIMILARITY.
FT   DISULFID   817   826       BY SIMILARITY.
FT   DISULFID   833   844       BY SIMILARITY.
FT   DISULFID   838   855       BY SIMILARITY.
```

```
FT   DISULFID    857    866       BY SIMILARITY.
FT   DISULFID    873    884       BY SIMILARITY.
FT   DISULFID    878    893       BY SIMILARITY.
FT   DISULFID    895    904       BY SIMILARITY.
FT   DISULFID    911    922       BY SIMILARITY.
FT   DISULFID    916    931       BY SIMILARITY.
FT   DISULFID    933    942       BY SIMILARITY.
FT   DISULFID    987    998       BY SIMILARITY.
FT   DISULFID    992   1007       BY SIMILARITY.
FT   DISULFID   1009   1018       BY SIMILARITY.
FT   DISULFID   1025   1036       BY SIMILARITY.
FT   DISULFID   1030   1045       BY SIMILARITY.
FT   DISULFID   1047   1056       BY SIMILARITY.
FT   DISULFID   1063   1074       BY SIMILARITY.
FT   DISULFID   1068   1083       BY SIMILARITY.
FT   DISULFID   1085   1094       BY SIMILARITY.
FT   DISULFID   1101   1122       BY SIMILARITY.
FT   DISULFID   1116   1131       BY SIMILARITY.
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.
FT   DISULFID   1225   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1271   1284       BY SIMILARITY.
FT   DISULFID   1276   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1311   1322       BY SIMILARITY.
FT   DISULFID   1316   1334       BY SIMILARITY.
FT   DISULFID   1336   1345       BY SIMILARITY.
FT   DISULFID   1352   1363       BY SIMILARITY.
FT   DISULFID   1357   1372       BY SIMILARITY.
FT   DISULFID   1374   1383       BY SIMILARITY.
FT   DISULFID   1391   1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           6.3%;  Score 708;  DB 1;  Length 2531;
  Best Local Similarity  29.1%;  Pred. No. 1.45e-121;
  Matches  161;  Conservative  116;  Mismatches 229;  Indels  48;  Gaps  38;

Db    420 ANPCEHAGKCL-NTLGSFECQCLQGYTGPRCEIDVNECISNPCQNDATC-L-D-QIGEFQ 475
          :||| : | |  : :  : | |  |: |  |:: :: |||||| : :|| | : :   |
Qy    920 SNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFW 979

Db    476 CICMPGYEGVYCEINTDECASSPCLHNGRCVDKINEFLCQCPKGFSGHLCQYDVDECA-S 534
          |||  |:||  ||:| |:|  : | :|: ||| ||:: | ||  ::| ||:  :| ||
Qy    980 CICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQD 1039

Db    535 -TPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHIGL-CKDGVATFTCLCQ 592
           || : :||:  |: : | || || | ||::|:|:|: : |  |  | |:|  :||:|
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    593 PGYTGHHCE-TN--I---:NECHSQPCRHGGTCQDRDNYYLCLCLKGTTGPNCEINLDDC 645
          ||:|  ||  :   :    : | :  |::|: |  | |  :| || |  | :|| :|
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE-KL--V 1156

Db    646 ASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPCHNGGTCEDGIAGFTCRCP 705
          : |  :  : |: | :     :   |  :  :  || :|    :|     ::  :  |
Qy   1157 SVNFINKESYLQ-IPSAKVRPQTNIT--L-QIATDEDSGILLYKGDKDHIAVELYRGRVR 1212

Db    706 EGYHDPTCLSEVNECNSNPCIHGACRD-GLNGYKCDCAPGWSGTNCDINNNEC-ESNPCV 763
          :| |    :    : :   | :  | :    : :  | | | |  :|
Qy   1213 ASY-DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNF 1271

Db    764 NGGTCKD-MTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTCIDDVAGYKCNCPLPY 821
          ::   ,  | : : | : |:: :| |  |::: |  |  :| :  ::|   :  : |:
Qy   1272 DSPLYVGGMPGKSNVASLRQA-PGQN-GTSFHGCIRNLYIN-SE-LQD---FQ-KVPMQ- 1322

Db    822 TGATCEVVLAPCATSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEIDINE-CVKSPCRH 880
          ||       ||    | : | |  | :  :|:| |  || |  |:  |: |: : | |
Qy   1323 TGILPGC--EPCHKKVCAH-GTCQPSSQ-AGFTCECQEGWMGPLCDQRTNDPCLGNKCVH 1378

Db    881 GASCQNTNG-SYRCLCQAGYTGRNCESDID--D-CRPNPCHNGGSCT-DGVNAAFCDCLP 935
          |: |   |: || | |  |  |  |: : |  : |::  | :|  |   |:  ::|:| :
Qy   1379 GT-CLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRLSGLGQPYCECSS 1436

Db    936 GFQGAFCEEDINEC 949
          |: |  |: :|: |
Qy   1437 GYTGDSCDREIS-C 1449


RESULT    7
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    44       POTENTIAL.
FT   CHAIN       45  2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45  1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746  1766       POTENTIAL.
FT   DOMAIN    1767  2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58  1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58    95       EGF-LIKE 1.
FT   DOMAIN      96   136       EGF-LIKE 2.
FT   DOMAIN     139   176       EGF-LIKE 3.
FT   DOMAIN     177   215       EGF-LIKE 4.
FT   DOMAIN     217   253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255   291       EGF-LIKE 6.
FT   DOMAIN     293   329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331   370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372   408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     409   447       EGF-LIKE 10.
FT   DOMAIN     449   486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488   524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526   562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564   600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602   637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639   675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677   713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715   751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753   789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791   827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829   865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     867   905       EGF-LIKE 22.
FT   DOMAIN     907   944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     946   982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984  1020       EGF-LIKE 25.
FT   DOMAIN    1022  1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1060  1096       EGF-LIKE 27.
FT   DOMAIN    1098  1134       EGF-LIKE 28.
FT   DOMAIN    1136  1181       EGF-LIKE 29.
FT   DOMAIN    1183  1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221  1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1259  1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1297  1335       EGF-LIKE 33.
FT   DOMAIN    1337  1373       EGF-LIKE 34.
FT   DOMAIN    1375  1412       EGF-LIKE 35.
FT   DOMAIN    1415  1451       EGF-LIKE 36.
FT   DOMAIN    1475  1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1475  1513       LIN/NOTCH 1.
FT   REPEAT    1514  1553       LIN/NOTCH 2.
FT   REPEAT    1554  1593       LIN/NOTCH 3.
FT   DOMAIN    1896  2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN    2538  2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID    62    73       BY SIMILARITY.
FT   DISULFID    67    83       BY SIMILARITY.
FT   DISULFID    85    94       BY SIMILARITY.
FT   DISULFID   100   111       BY SIMILARITY.
FT   DISULFID   105   124       BY SIMILARITY.
FT   DISULFID   126   135       BY SIMILARITY.
FT   DISULFID   143   154       BY SIMILARITY.
FT   DISULFID   148   164       BY SIMILARITY.
FT   DISULFID   166   175       BY SIMILARITY.
FT   DISULFID   181   192       BY SIMILARITY.
FT   DISULFID   186   203       BY SIMILARITY.
FT   DISULFID   205   214       BY SIMILARITY.
FT   DISULFID   221   232       BY SIMILARITY.
FT   DISULFID   226   241       BY SIMILARITY.
FT   DISULFID   243   252       BY SIMILARITY.
FT   DISULFID   259   270       BY SIMILARITY.
FT   DISULFID   264   279       BY SIMILARITY.
FT   DISULFID   281   290       BY SIMILARITY.
FT   DISULFID   297   308       BY SIMILARITY.
FT   DISULFID   302   317       BY SIMILARITY.
FT   DISULFID   319   328       BY SIMILARITY.
FT   DISULFID   335   349       BY SIMILARITY.
FT   DISULFID   343   358       BY SIMILARITY.
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   376   387       BY SIMILARITY.
FT   DISULFID   381   396       BY SIMILARITY.
FT   DISULFID   398   407       BY SIMILARITY.
FT   DISULFID   413   424       BY SIMILARITY.
FT   DISULFID   418   435       BY SIMILARITY.
FT   DISULFID   437   446       BY SIMILARITY.
FT   DISULFID   453   465       BY SIMILARITY.
FT   DISULFID   459   474       BY SIMILARITY.
FT   DISULFID   476   485       BY SIMILARITY.
FT   DISULFID   492   503       BY SIMILARITY.
FT   DISULFID   497   512       BY SIMILARITY.
FT   DISULFID   514   523       BY SIMILARITY.
FT   DISULFID   530   541       BY SIMILARITY.
FT   DISULFID   535   550       BY SIMILARITY.
FT   DISULFID   552   561       BY SIMILARITY.
FT   DISULFID   568   579       BY SIMILARITY.
FT   DISULFID   573   588       BY SIMILARITY.
FT   DISULFID   590   599       BY SIMILARITY.
FT   DISULFID   606   616       BY SIMILARITY.
FT   DISULFID   611   625       BY SIMILARITY.
FT   DISULFID   627   636       BY SIMILARITY.
FT   DISULFID   643   654       BY SIMILARITY.
FT   DISULFID   648   663       BY SIMILARITY.
FT   DISULFID   665   674       BY SIMILARITY.
FT   DISULFID   681   692       BY SIMILARITY.
FT   DISULFID   686   701       BY SIMILARITY.
FT   DISULFID   703   712       BY SIMILARITY.
FT   DISULFID   719   730       BY SIMILARITY.
FT   DISULFID   724   739       BY SIMILARITY.
FT   DISULFID   741   750       BY SIMILARITY.
FT   DISULFID   757   768       BY SIMILARITY.
FT   DISULFID   762   777       BY SIMILARITY.
FT   DISULFID   779   788       BY SIMILARITY.
FT   DISULFID   795   806       BY SIMILARITY.
FT   DISULFID   800   815       BY SIMILARITY.
FT   DISULFID   817   826       BY SIMILARITY.
FT   DISULFID   833   844       BY SIMILARITY.
FT   DISULFID   838   853       BY SIMILARITY.
FT   DISULFID   855   864       BY SIMILARITY.
...
Note: remainder of annotations omitted.
```

Query Match 6.3%; Score 708; DB 1; Length 2703;
Best Local Similarity 44.8%; Pred. No. 1.45e-121;

```
  Matches  108;  Conservative   37;  Mismatches  82;  Indels  14;  Gaps   7;

Db    490 NECESHPCQNEGSCLDDPGTF-RCVCMPGFTGTQCEIDIDECQSNPCLNDGTCH--D-KI 545
          | | |:|| |:|:|  ||  | || |  || | :|:: |  | |||| : ||||  :
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    546 NGFKCSCALGFTGARCQINIDDCQSQPCRNRGICHDSIAGYSCECPPGYTGTSCEININD 605
          :|| | || || |  |::|:|||:   | | : | |:|  |:| ||| |||  || :::
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    606 C--DSNPCHR-GKCIDDVNSFKCLCDPGYTGYICQKQINECESNPCQFDGHCQDRVGSYY 662
          |  | |||:: :|||   ::||| | ||| |  |: ::::|: | |   :|| | | :|
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    663 CQCQAGTSGKNCE-----V--NVNECHSNPCNNGATCIDGINSYKCQCVPGFTGQHCEKN 715
          | |  | ||  ||     |    : | :  | ||| ||  ||   |||:||: |: |||
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155

Db    716 V 716
          |
      1156 V 1156


RESULT    8
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2075       ANK MOTIF 6.
FT   CARBOHYD   888    888       POTENTIAL.
FT   CARBOHYD   959    959       POTENTIAL.
FT   CARBOHYD  1179   1179       POTENTIAL.
FT   CARBOHYD  1241   1241       POTENTIAL.
FT   CARBOHYD  1489   1489       POTENTIAL.
FT   CARBOHYD  1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match           6.2%;  Score 704;  DB 1;  Length 2531;
  Best Local Similarity  28.5%;  Pred. No. 1.14e-120;
  Matches  158;  Conservative  118;  Mismatches  230;  Indels  48;  Gaps  38;

Db    420 ANRCEHAGKCL-NTLGSFECQCLQGYTGPGCEIDVNECISNPCQNDATC-L-D-QIGEFQ 475
          :| | : | |  : :  : | |  |: |  |:: :: |||||| : :|| | : :   |
Qy    920 SNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFW 979

Db    476 CICMPGYEGVYCEINTDECASSPCLHNGHCMDKIHEFQCQCPKGFNGHLCQYDVDECA-S 534
          |||  |:||  ||:| |:|  : | :|: |:| |::: | ||  : | ||:  :| ||
Qy    980 CICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQD 1039

Db    535 -TPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGS-CKDGVATFTCLCQ 592
           || : :||:  |: : | || || | ||::|:|:|: : |  |: | |:|  :||:|
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    593 PGYTGHHCE-TN--I----NECHSQPCRHGGTCQDRDNSYLCLCLKGTTGPNCEINLDDC 645
          ||:|  || :   :    : | :  |::|: |  | |  :| || |  | :|| :|
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE-KL--V 1156

Db    646 ASNPCDSGTCLDKIDGYECACEPGYTGSMCNVVNIDECAGSPCHNGGTCEDGIAGFTCRCP 705
          : |  :  : |: | :     :   |  :  :  || :|    :|     ::  :  |
Qy   1157 SVNFINKESYLQ-IPSAKVRPQTNIT--L-QIATDEDSGILLYKGDKDHIAVELYRGRVR 1212

Db    706 EGYHDPTCLSEVNECNSNPCIHGACRD-GLNGYKCDCAPGWSGTNCDINNNEC-ESNPCV 763
          :| |    :    : :    |  :  | :    : :  | | |  |  |  :|
Qy   1213 ASY-DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNF 1271

Db    764 NGGTCKD-MTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTCIDDVAGYKCNCPLPY 821
          ::      | : : | : |:: :| |  |::: |  |  :| :  ::|   :  : |:
Qy   1272 DSPLYVGGMPGKSNVASLRQA-PGQN-GTSFHGCIRNLYIN-SE-LQD---FQ-KVPMQ- 1322

Db    822 TGATCEVVLAPCATSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEVDINE-CVKSPCRH 880
          ||        ||    | : | |  | :  :|:| |  || |  |:   |: |: : | |
Qy   1323 TGILPGC--EPCHKKVCAH-GTCQPSSQ-AGFTCECQEGWMGPLCDQRTNDPCLGNKCVH 1378

Db    881 GASCQNTNG-SYRCLCQAGYTGRNCESDID--D-CRPNPCHNGGSCT-DGINTAFCDCLP 935
          |: |   |: || | |  |  |  |: : |  : |::  | :|  |   |:  ::|:| :
Qy   1379 GT-CLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRLSGLGQPYCECSS 1436
```

```
Db    936 GFQGAFCEEDINEC 949
          |: |  |: :|: |
Qy   1437 GYTGDSCDREIS-C 1449


RESULT    9
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21   1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725   1747       POTENTIAL.
FT   DOMAIN    1748   2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     57       EGF-LIKE 1.
FT   DOMAIN      58     98       EGF-LIKE 2.
FT   DOMAIN     101    138       EGF-LIKE 3.
FT   DOMAIN     139    175       EGF-LIKE 4.
FT   DOMAIN     177    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     827    865       EGF-LIKE 22.
FT   DOMAIN     867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     981   1017       EGF-LIKE 26.
FT   DOMAIN    1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1057   1093       EGF-LIKE 28.
FT   DOMAIN    1095   1141       EGF-LIKE 29.
FT   DOMAIN    1143   1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1181   1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1219   1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1265   1303       EGF-LIKE 33.
FT   DOMAIN    1305   1344       EGF-LIKE 34.
FT   DOMAIN    1346   1382       EGF-LIKE 35.
FT   DOMAIN    1385   1423       EGF-LIKE 36.
FT   DOMAIN    1446   1561       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446   1486       LIN/NOTCH 1.
FT   REPEAT    1487   1520       LIN/NOTCH 2.
FT   REPEAT    1521   1561       LIN/NOTCH 3.
FT   DOMAIN    1861   2074       6 X ANK MOTIF REPEATS.
FT   REPEAT    1861   1891       ANK MOTIF 1.
FT   REPEAT    1892   1940       ANK MOTIF 1.
FT   REPEAT    1941   1974       ANK MOTIF 1.
FT   REPEAT    1975   2007       ANK MOTIF 1.
FT   REPEAT    2008   2040       ANK MOTIF 1.
FT   REPEAT    2041   2074       ANK MOTIF 1.
FT   DOMAIN    2265   2276       POLY-GLN (OPA-REPEAT).
FT   DISULFID    25     35       BY SIMILARITY.
FT   DISULFID    29     45       BY SIMILARITY.
FT   DISULFID    47     56       BY SIMILARITY.
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     86       BY SIMILARITY.
FT   DISULFID    88     97       BY SIMILARITY.
FT   DISULFID   105    116       BY SIMILARITY.
FT   DISULFID   110    126       BY SIMILARITY.
FT   DISULFID   128    137       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    163       BY SIMILARITY.
FT   DISULFID   165    174       BY SIMILARITY.
FT   DISULFID   181    194       BY SIMILARITY.
FT   DISULFID   188    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    242       BY SIMILARITY.
FT   DISULFID   244    253       BY SIMILARITY.
FT   DISULFID   260    271       BY SIMILARITY.
FT   DISULFID   265    280       BY SIMILARITY.
FT   DISULFID   282    291       BY SIMILARITY.
FT   DISULFID   298    311       BY SIMILARITY.
FT   DISULFID   305    320       BY SIMILARITY.
FT   DISULFID   322    331       BY SIMILARITY.
FT   DISULFID   338    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   375    386       BY SIMILARITY.
FT   DISULFID   380    397       BY SIMILARITY.
FT   DISULFID   399    408       BY SIMILARITY.
FT   DISULFID   415    428       BY SIMILARITY.
FT   DISULFID   422    437       BY SIMILARITY.
FT   DISULFID   439    448       BY SIMILARITY.
```

```
FT   DISULFID    455    466       BY SIMILARITY.
FT   DISULFID    460    475       BY SIMILARITY.
FT   DISULFID    477    486       BY SIMILARITY.
FT   DISULFID    493    503       BY SIMILARITY.
FT   DISULFID    498    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    578       BY SIMILARITY.
FT   DISULFID    573    587       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    605    616       BY SIMILARITY.
FT   DISULFID    610    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    653       BY SIMILARITY.
FT   DISULFID    648    662       BY SIMILARITY.
FT   DISULFID    664    673       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   DISULFID    685    700       BY SIMILARITY.
FT   DISULFID    702    711       BY SIMILARITY.
FT   DISULFID    718    728       BY SIMILARITY.
FT   DISULFID    723    737       BY SIMILARITY.
FT   DISULFID    739    748       BY SIMILARITY.
FT   DISULFID    755    766       BY SIMILARITY.
FT   DISULFID    760    775       BY SIMILARITY.
FT   DISULFID    777    786       BY SIMILARITY.
FT   DISULFID    793    804       BY SIMILARITY.
FT   DISULFID    798    813       BY SIMILARITY.
FT   DISULFID    815    824       BY SIMILARITY.
FT   DISULFID    831    842       BY SIMILARITY.
FT   DISULFID    836    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
FT   DISULFID    871    882       BY SIMILARITY.
FT   DISULFID    876    891       BY SIMILARITY.
FT   DISULFID    893    902       BY SIMILARITY.
FT   DISULFID    909    920       BY SIMILARITY.
FT   DISULFID    914    929       BY SIMILARITY.
FT   DISULFID    931    940       BY SIMILARITY.
FT   DISULFID    947    958       BY SIMILARITY.
FT   DISULFID    952    967       BY SIMILARITY.
FT   DISULFID    969    978       BY SIMILARITY.
FT   DISULFID   1023   1034       BY SIMILARITY.
FT   DISULFID   1028   1043       BY SIMILARITY.
FT   DISULFID   1045   1054       BY SIMILARITY.
FT   DISULFID   1061   1072       BY SIMILARITY.
FT   DISULFID   1066   1081       BY SIMILARITY.
FT   DISULFID   1083   1092       BY SIMILARITY.
FT   DISULFID   1099   1120       BY SIMILARITY.
FT   DISULFID   1114   1129       BY SIMILARITY.
FT   DISULFID   1131   1140       BY SIMILARITY.
FT   DISULFID   1147   1158       BY SIMILARITY.
FT   DISULFID   1152   1167       BY SIMILARITY.
FT   DISULFID   1169   1178       BY SIMILARITY.
FT   DISULFID   1185   1196       BY SIMILARITY.
FT   DISULFID   1190   1205       BY SIMILARITY.
FT   DISULFID   1207   1216       BY SIMILARITY.
FT   DISULFID   1223   1242       BY SIMILARITY.
FT   DISULFID   1236   1251       BY SIMILARITY.
FT   DISULFID   1253   1262       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           5.9%;  Score 667;  DB 1;  Length 2437;
  Best Local Similarity  39.3%;  Pred. No. 2.16e-112;
  Matches   94;  Conservative   53;  Mismatches  79;  Indels  13;  Gaps   9;

Db    791 NECASNPCLNQGSCIDDVAGF-KCNCMLPYTGEVCENVLAPCSPRPCKNGGVC--RESED 847
          | | |||| |:|:|  |   | :| |   : |: |:  : :| : |||:|| |  :|:|:
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    848 FQSFSCNCPAGWQGQTCEVDINECVRNPCTNGGVCENLRGGFQCRCNPGFTGALCENDID 907
          ::| | |: | :|: |||::::|  | | | : | :    : | | | :|| |||: :|

Qy    976 -DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    908 DC--EPNPCSNGGVCQDRVNGFVCVCLAGFRGERCAEDIDECVSAPCRNGGNCTDCVNSY 965
          |  : ||| : : |    :|| | | :|: ||:|  |:|:|    |:||::||| ||:|
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    966 TCSCPAGFSGINCEINTPDC-TESS-C--FN--GGT-CVDGISSFSCVCLPGFTGNYCQ 1017
          || || |:||: ||:: |     :| |  |:   |: |:  |:   | ||||: |: |:
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT  10
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   >2444      NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN       19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1737   1757       POTENTIAL.
FT   DOMAIN     1758   >2444      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20     58       EGF-LIKE 1.
FT   DOMAIN       59     99       EGF-LIKE 2.
FT   DOMAIN      102    139       EGF-LIKE 3.
FT   DOMAIN      140    176       EGF-LIKE 4.
FT   DOMAIN      178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      218    255       EGF-LIKE 6.
FT   DOMAIN      257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372    410       EGF-LIKE 10.
FT   DOMAIN      412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789       EGF-LIKE 20.
FT   DOMAIN      791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    868       EGF-LIKE 22.
FT   DOMAIN      870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      908    944       EGF-LIKE 24.
FT   DOMAIN      946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984   1020       EGF-LIKE 26.
FT   DOMAIN     1022   1058       EGF-LIKE 27.
FT   DOMAIN     1060   1096       EGF-LIKE 28.
FT   DOMAIN     1098   1144       EGF-LIKE 29.
FT   DOMAIN     1146   1182       EGF-LIKE 30.
FT   DOMAIN     1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1268   1306       EGF-LIKE 33.
FT   DOMAIN     1308   1347       EGF-LIKE 34.
FT   DOMAIN     1349   1385       EGF-LIKE 35.
FT   DOMAIN     1388   1427       EGF-LIKE 36.
FT   DOMAIN     1446   1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1481       LIN/NOTCH 1.
FT   REPEAT     1482   1523       LIN/NOTCH 2.
FT   REPEAT     1524   1563       LIN/NOTCH 3.
FT   DOMAIN     1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT     1876   1921       ANK MOTIF 1.
FT   REPEAT     1923   1954       ANK MOTIF 2.
FT   REPEAT     1956   1987       ANK MOTIF 3.
FT   REPEAT     1990   2021       ANK MOTIF 4.
FT   REPEAT     2023   2054       ANK MOTIF 5.
FT   REPEAT     2056   2087       ANK MOTIF 6.
FT   DOMAIN     1576   1579       POLY-VAL.
FT   DOMAIN     1662   1665       POLY-ARG.
FT   DOMAIN     1729   1732       POLY-PRO.
FT   DOMAIN     1741   1744       POLY-ALA.
FT   DOMAIN     1902   1905       POLY-GLU.
FT   DOMAIN     2260   2263       POLY-GLY.
FT   DOMAIN     2404   2407       POLY-GLN.
FT   DOMAIN     2411   2418       POLY-PRO.
FT   DISULFID     24     37       BY SIMILARITY.
FT   DISULFID     31     46       BY SIMILARITY.
FT   DISULFID     48     57       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    261    272       BY SIMILARITY.
FT   DISULFID    266    281       BY SIMILARITY.
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    299    312       BY SIMILARITY.
FT   DISULFID    306    321       BY SIMILARITY.
FT   DISULFID    323    332       BY SIMILARITY.
FT   DISULFID    339    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    361    370       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    398       BY SIMILARITY.
FT   DISULFID    400    409       BY SIMILARITY.
FT   DISULFID    416    429       BY SIMILARITY.
FT   DISULFID    423    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    456    467       BY SIMILARITY.
FT   DISULFID    461    476       BY SIMILARITY.
FT   DISULFID    478    487       BY SIMILARITY.
FT   DISULFID    494    505       BY SIMILARITY.
FT   DISULFID    499    514       BY SIMILARITY.
FT   DISULFID    516    525       BY SIMILARITY.
FT   DISULFID    532    543       BY SIMILARITY.
FT   DISULFID    537    552       BY SIMILARITY.
FT   DISULFID    554    563       BY SIMILARITY.
FT   DISULFID    570    580       BY SIMILARITY.
FT   DISULFID    575    589       BY SIMILARITY.
FT   DISULFID    591    600       BY SIMILARITY.
FT   DISULFID    607    618       BY SIMILARITY.
FT   DISULFID    612    627       BY SIMILARITY.
FT   DISULFID    629    638       BY SIMILARITY.
FT   DISULFID    645    655       BY SIMILARITY.
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    867       BY SIMILARITY.
FT   DISULFID    874    885       BY SIMILARITY.
FT   DISULFID    879    894       BY SIMILARITY.
FT   DISULFID    896    905       BY SIMILARITY.
FT   DISULFID    912    923       BY SIMILARITY.
FT   DISULFID    917    932       BY SIMILARITY.
FT   DISULFID    934    943       BY SIMILARITY.
FT   DISULFID    988    999       BY SIMILARITY.
FT   DISULFID    993   1008       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1026   1037       BY SIMILARITY.
FT   DISULFID   1031   1046       BY SIMILARITY.
FT   DISULFID   1048   1057       BY SIMILARITY.
FT   DISULFID   1064   1075       BY SIMILARITY.
FT   DISULFID   1069   1084       BY SIMILARITY.
FT   DISULFID   1086   1095       BY SIMILARITY.
FT   DISULFID   1102   1123       BY SIMILARITY.
FT   DISULFID   1117   1132       BY SIMILARITY.
FT   DISULFID   1134   1143       BY SIMILARITY.
FT   DISULFID   1150   1161       BY SIMILARITY.
FT   DISULFID   1155   1170       BY SIMILARITY.
FT   DISULFID   1172   1181       BY SIMILARITY.
FT   DISULFID   1188   1199       BY SIMILARITY.
FT   DISULFID   1193   1208       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           5.9%;  Score 667;  DB 1;  Length 2444;
  Best Local Similarity  30.2%;  Pred. No. 2.16e-112;
  Matches  167;  Conservative  115;  Mismatches  220;  Indels  51;  Gaps  37;
```

```
Db    832 PCAPSPCRNGGECRQSEDYESFSCVCPTAGAKGQTCEVDINECVLSPCRHGASCQNTHGX 891
Qy    917 PCLSNPCKNDGTC-NSDPVDFYRCTCPY-GFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    892 ---YRCHCQAGYSGRNCETDIDDCRPNPCHNGGSCTDGINTAFCDCLPGFRGTFCEEDIN 948
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    949 ECASD--PCRNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSF 1006
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db   1007 TCLCPPGFTGSYCQ---HVV----NECDSRPCLLGGTCQDGRGLHRCTCPQGYTGPNCQN 1059
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEK 1154

Db   1060 LV--HWCDSSPC-K-NGGKCW-QTH-T-QYRCECPSGWTGLY-CDVPSVSCEV-AAQ-RQ 1109
Qy   1155 LVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSG-ILLYKGDKDHIAVELYRGRVRA 1213

Db   1110 GVDVARLCQHGGLCVDAGNTHHCRCQAGYTGSYCEDL-VDECSPSPCQNGATCTDYLGGY 1168
Qy   1214 SYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSK-QSTLNFD 1272

Db   1169 SCKCVAGYHGVNCSEEIDECLSH-PCQNGGTCLD-LPNTYKCSCPRGTQGVHCEINVDDC 1226
Qy   1273 SPLYVGGMPG-K-SN-VAS-LRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGI--- 1325

Db   1227 NPPVDPVSRSPKCFNNGTC-VDQVGGYSCTCPPGFVGERCEGDVNE-CLSNPCDARGTQN 1284
Qy   1326 LPGCEPCHKKV-C-AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKC-VHGT-- 1380

Db   1285 CVQRVN-DFHCECRAGHTGRRC---ESVINGCKGKPCKNGGTCAVASNTARGFICKCPAG 1340
Qy   1381 CLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRL-SGLGQPY-CECSSG 1437

Db   1341 FEGATCENDARTC 1353
Qy   1438 YTGDSCDREI-SC 1449


RESULT  11
ID   DLL1_HUMAN     STANDARD;      PRT;   723 AA.
AC   O00548;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MANN R.S., GRAY G.E., HENRIQUE D., ISH-HOROWICZ D.,
RA   ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF003522; G2197069; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   723       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   723       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     226   254       EGF-LIKE 1.
FT   DOMAIN     257   285       EGF-LIKE 2.
FT   DOMAIN     292   325       EGF-LIKE 3.
FT   DOMAIN     332   363       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     370   402       EGF-LIKE 5.
FT   DOMAIN     409   440       EGF-LIKE 6.
FT   DOMAIN     447   478       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     485   516       EGF-LIKE 8.
FT   DISULFID   226   237       BY SIMILARITY.
FT   DISULFID   230   243       BY SIMILARITY.
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   257   268       BY SIMILARITY.
FT   DISULFID   263   274       BY SIMILARITY.
FT   DISULFID   276   285       BY SIMILARITY.
FT   DISULFID   292   304       BY SIMILARITY.
FT   DISULFID   298   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   391       BY SIMILARITY.
FT   DISULFID   393   402       BY SIMILARITY.
FT   DISULFID   409   420       BY SIMILARITY.
FT   DISULFID   414   429       BY SIMILARITY.
FT   DISULFID   431   440       BY SIMILARITY.
FT   DISULFID   447   467       BY SIMILARITY.
FT   DISULFID   469   478       BY SIMILARITY.
FT   DISULFID   485   496       BY SIMILARITY.
FT   DISULFID   490   505       BY SIMILARITY.
FT   DISULFID   507   516       BY SIMILARITY.
FT   CARBOHYD   477   477       POTENTIAL.
SQ   SEQUENCE   723 AA;  77956 MW;  A1D48BDB CRC32;

  Query Match           5.7%;  Score 646;  DB 1;  Length 723;
  Best Local Similarity  38.2%;  Pred. No. 1.03e-107;
  Matches   91;  Conservative   53;  Mismatches  81;  Indels  13;  Gaps   9;

Db    296 KPCKNGATCTNTGQGSYTCSCRPGYTGATCELGIDECDPSPCKNGGSC--TD-LENSYSC 352
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    353 TCPPGFYGKICELSAMTCADGPCFNGGRCSDSPDGGYSCRCPVGYSGFNCEKKIDYCS-S 411
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    412 -SPCSNGAKCVDLGDAYLCRCQAGFSGRHCDDNVDDCASSPCANGGTCRDGVNDFSCTCP 470
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    471 PGYTGRNC--SAP-V-SR---CEHAPCHNGATCHERGHGYVCECARGYGGPNCQFLLP 521
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT  12
ID   DLL1_RAT       STANDARD;      PRT;   714 AA.
AC   P97677;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
```

```
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DISIBIO G., HEBSHI L., BOULTER J., WEINMASTER G.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78889; G1699046; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   714       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   537       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   538   560       POTENTIAL.
FT   DOMAIN     561   714       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225   253       EGF-LIKE 1.
FT   DOMAIN     256   284       EGF-LIKE 2.
FT   DOMAIN     291   324       EGF-LIKE 3.
FT   DOMAIN     331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369   401       EGF-LIKE 5.
FT   DOMAIN     408   439       EGF-LIKE 6.
FT   DOMAIN     446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484   515       EGF-LIKE 8.
FT   DISULFID   225   236       BY SIMILARITY.
FT   DISULFID   229   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   262   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   303       BY SIMILARITY.
FT   DISULFID   297   313       BY SIMILARITY.
FT   DISULFID   315   324       BY SIMILARITY.
FT   DISULFID   331   342       BY SIMILARITY.
FT   DISULFID   336   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   380       BY SIMILARITY.
FT   DISULFID   374   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   CARBOHYD   476   476       POTENTIAL.
SQ   SEQUENCE   714 AA;  77378 MW;  604B76D1 CRC32;
```

```
  Query Match           5.6%;  Score 631;  DB 1;  Length 714;
  Best Local Similarity  36.6%;  Pred. No. 2.22e-104;
  Matches   87;  Conservative   56;  Mismatches  82;  Indels  13;  Gaps   9;

Db    295 KPCRNGATCTNTGQGSYTCSCRPGYTGANCELEVDECAPSPCRNGGSC--TD-LEDSYSC 351
          :||:| :|| :    | |:|  |: | :|:: :  | ::||::||:|   :  ||:: |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    352 TCPPGFYGKVCELSAMTCADGPCFNGGRCSDNPDGGYTCHCPAGFSGFNCEKKIDLCS-S 410
          |: || |  ||::   | |  | | : | |  :  ||| ||: ::|  || |:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    411 -SPCSNGAKCVDLGNSYLCRCQTGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCP 469
          :|| : :||:   ::: | |  |: |  |: : |||  : | ||: | |:|| ::| ||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    470 PGYTGRNC--SAP-V-SR---CEHAPCHNGATCHQRGQRYMCECAQGYGGANCQFLLP 520
          ||:|  |  |:| | :|   |::  |:||| |  |    :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT  13
ID   DLL1_MOUSE     STANDARD;      PRT;   722 AA.
AC   Q61483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C X C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95401858.
RA   BETTENHAUSEN B., DE ANGELIS M.H., SIMON D., GUENET J.-L., GOSSLER A.;
RT   "Transient and restricted expression during mouse embryogenesis of
RT   Dll1, a murine gene closely related to Drosophila Delta.";
RL   DEVELOPMENT 121:2407-2418(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN THE PARAXIAL
CC       MESODERM AND NERVOUS SYSTEM. EXPRESSED AT HIGH LEVELS IN ADULT
CC       HEART AND AT LOWER LEVELS, IN ADULT LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED UNTIL DAY 15 IN THE EMBRYO.
CC       EXPRESSION THEN DECREASES AND INCREASES AGAIN IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80903; G806570; -.
DR   MGD; MGI:104659; DLL1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   722       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
```

```
FT   DOMAIN      569    722       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      225    253       EGF-LIKE 1.
FT   DOMAIN      256    284       EGF-LIKE 2.
FT   DOMAIN      291    324       EGF-LIKE 3.
FT   DOMAIN      331    362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN      369    401       EGF-LIKE 5.
FT   DOMAIN      408    439       EGF-LIKE 6.
FT   DOMAIN      446    477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      484    515       EGF-LIKE 8.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    229    242       BY SIMILARITY.
FT   DISULFID    244    253       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    262    273       BY SIMILARITY.
FT   DISULFID    275    284       BY SIMILARITY.
FT   DISULFID    291    303       BY SIMILARITY.
FT   DISULFID    297    313       BY SIMILARITY.
FT   DISULFID    315    324       BY SIMILARITY.
FT   DISULFID    331    342       BY SIMILARITY.
FT   DISULFID    336    351       BY SIMILARITY.
FT   DISULFID    353    362       BY SIMILARITY.
FT   DISULFID    369    380       BY SIMILARITY.
FT   DISULFID    374    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   DISULFID    446    466       BY SIMILARITY.
FT   DISULFID    468    477       BY SIMILARITY.
FT   DISULFID    484    495       BY SIMILARITY.
FT   DISULFID    489    504       BY SIMILARITY.
FT   DISULFID    506    515       BY SIMILARITY.
FT   CARBOHYD    476    476       POTENTIAL.
SQ   SEQUENCE   722 AA;  78448 MW;  5A647702 CRC32;

  Query Match           5.5%;  Score 627;  DB 1;  Length 722;
  Best Local Similarity  37.0%;  Pred. No. 1.71e-103;
  Matches   88;  Conservative   55;  Mismatches  82;  Indels  13;  Gaps   9;

Db    295 KPCRNGATCTNTGQGSYTCSCRPGYTGANCELEVDECAPSPCKNGASC--TD-LEDSFSC 351
          :||:| :|| :     | |:|  |: | :|:: :  | ::|||:|::|   :  ||:| |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    352 TCPPGFYGKVCELSAMTCADGPCFNGGRCSDNPDGGYTCHCPLGFSGFNCEKKMDLCG-S 410
          |: || |  ||::   | |  | | : | |  :  ||| ||  ::|  || |:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    411 -SPCSNGAKCVDLGNSYLCRCQAGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCP 469
           :|| : :||:   ::: | | :|: |  |: : |||  : | ||: | |:|| ::| ||
      1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    470 PGYTGKNC--SAP-V-SR---CEHAPCHNGATCHQRGQRYMCECAQGYGGPNCQFLLP 520
          ||:|  |  |:| | :|   |::  |:||| |  |    :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT   14
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. VIROL. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG, AND TESTIS;
RX   MEDLINE; 96281668.
RA   UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   DEVELOPMENT 122:2251-2259(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; G1714084; -.
DR   EMBL; U43691; G1401160; -.
DR   PIR; A38072; TVMVT3.
DR   MGD; MGI:107471; NOTCH4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   DOMAIN      21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1444   1464       POTENTIAL.
FT   DOMAIN    1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     60       EGF-LIKE 1.
FT   DOMAIN      61    112       EGF-LIKE 2.
FT   DOMAIN     115    152       EGF-LIKE 3.
FT   DOMAIN     153    189       EGF-LIKE 4.
FT   DOMAIN     191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     231    271       EGF-LIKE 6.
FT   DOMAIN     273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     389    427       EGF-LIKE 10.
FT   DOMAIN     429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     622    656       EGF-LIKE 16.
FT   DOMAIN     658    686       EGF-LIKE 17.
FT   DOMAIN     688    724       EGF-LIKE 18.
FT   DOMAIN     726    762       EGF-LIKE 19.
FT   DOMAIN     764    800       EGF-LIKE 20.
FT   DOMAIN     803    839       EGF-LIKE 21.
FT   DOMAIN     841    877       EGF-LIKE 22.
FT   DOMAIN     878    924       EGF-LIKE 23.
FT   DOMAIN     926    962       EGF-LIKE 24.
FT   DOMAIN     964   1000       EGF-LIKE 25.
```

```
FT   DOMAIN     1002   1040      EGF-LIKE 26.
FT   DOMAIN     1042   1081      EGF-LIKE 27.
FT   DOMAIN     1083   1122      EGF-LIKE 28.
FT   DOMAIN     1126   1167      EGF-LIKE 29.
FT   DOMAIN     1168   1282      3 X LIN/NOTCH REPEATS.
FT   REPEAT     1168   1208      LIN/NOTCH 1.
FT   REPEAT     1209   1242      LIN/NOTCH 2.
FT   REPEAT     1243   1282      LIN/NOTCH 3.
FT   DOMAIN     1572   1785      6 X ANK MOTIF REPEATS.
FT   REPEAT     1572   1603      ANK MOTIF 1.
FT   REPEAT     1622   1653      ANK MOTIF 2.
FT   REPEAT     1654   1685      ANK MOTIF 3.
FT   REPEAT     1688   1719      ANK MOTIF 4.
FT   REPEAT     1721   1752      ANK MOTIF 5.
FT   REPEAT     1754   1785      ANK MOTIF 6.
FT   DISULFID     25     38      BY SIMILARITY.
FT   DISULFID     32     48      BY SIMILARITY.
FT   DISULFID     50     59      BY SIMILARITY.
FT   DISULFID     65     77      BY SIMILARITY.
FT   DISULFID     71    100      BY SIMILARITY.
FT   DISULFID    102    111      BY SIMILARITY.
FT   DISULFID    119    130      BY SIMILARITY.
FT   DISULFID    124    140      BY SIMILARITY.
FT   DISULFID    142    151      BY SIMILARITY.
FT   DISULFID    157    168      BY SIMILARITY.
FT   DISULFID    162    177      BY SIMILARITY.
FT   DISULFID    179    188      BY SIMILARITY.
FT   DISULFID    195    208      BY SIMILARITY.
FT   DISULFID    202    217      BY SIMILARITY.
FT   DISULFID    219    228      BY SIMILARITY.
FT   DISULFID    235    246      BY SIMILARITY.
FT   DISULFID    240    259      BY SIMILARITY.
FT   DISULFID    261    270      BY SIMILARITY.
FT   DISULFID    235    246      BY SIMILARITY.
FT   DISULFID    240    259      BY SIMILARITY.
FT   DISULFID    261    270      BY SIMILARITY.
FT   DISULFID    277    288      BY SIMILARITY.
FT   DISULFID    282    297      BY SIMILARITY.
FT   DISULFID    299    308      BY SIMILARITY.
FT   DISULFID    315    329      BY SIMILARITY.
FT   DISULFID    323    338      BY SIMILARITY.
FT   DISULFID    340    349      BY SIMILARITY.
FT   DISULFID    356    367      BY SIMILARITY.
FT   DISULFID    361    376      BY SIMILARITY.
FT   DISULFID    378    387      BY SIMILARITY.
FT   DISULFID    393    404      BY SIMILARITY.
FT   DISULFID    398    415      BY SIMILARITY.
FT   DISULFID    417    426      BY SIMILARITY.
FT   DISULFID    433    449      BY SIMILARITY.
FT   DISULFID    443    458      BY SIMILARITY.
FT   DISULFID    460    469      BY SIMILARITY.
FT   DISULFID    476    487      BY SIMILARITY.
FT   DISULFID    481    496      BY SIMILARITY.
FT   DISULFID    498    507      BY SIMILARITY.
FT   DISULFID    514    525      BY SIMILARITY.
FT   DISULFID    519    534      BY SIMILARITY.
FT   DISULFID    536    545      BY SIMILARITY.
FT   DISULFID    552    563      BY SIMILARITY.
FT   DISULFID    557    572      BY SIMILARITY.
FT   DISULFID    574    583      BY SIMILARITY.
FT   DISULFID    590    601      BY SIMILARITY.
FT   DISULFID    595    610      BY SIMILARITY.
FT   DISULFID    612    621      BY SIMILARITY.
FT   DISULFID    626    637      BY SIMILARITY.
FT   DISULFID    631    646      BY SIMILARITY.
FT   DISULFID    648    655      BY SIMILARITY.
FT   DISULFID    662    669      BY SIMILARITY.
FT   DISULFID    664    674      BY SIMILARITY.
FT   DISULFID    676    685      BY SIMILARITY.
FT   DISULFID    692    703      BY SIMILARITY.
FT   DISULFID    697    712      BY SIMILARITY.
FT   DISULFID    714    723      BY SIMILARITY.
FT   DISULFID    730    741      BY SIMILARITY.
FT   DISULFID    735    750      BY SIMILARITY.
FT   DISULFID    752    761      BY SIMILARITY.
FT   DISULFID    768    779      BY SIMILARITY.
FT   DISULFID    773    788      BY SIMILARITY.
FT   DISULFID    790    799      BY SIMILARITY.
FT   DISULFID    807    818      BY SIMILARITY.
FT   DISULFID    812    827      BY SIMILARITY.
FT   DISULFID    829    838      BY SIMILARITY.
FT   DISULFID    845    856      BY SIMILARITY.
FT   DISULFID    850    865      BY SIMILARITY.
FT   DISULFID    867    876      BY SIMILARITY.
FT   DISULFID    882    903      BY SIMILARITY.
FT   DISULFID    897    912      BY SIMILARITY.
FT   DISULFID    914    923      BY SIMILARITY.
FT   DISULFID    930    941      BY SIMILARITY.
FT   DISULFID    935    950      BY SIMILARITY.
FT   DISULFID    952    961      BY SIMILARITY.
FT   DISULFID    968    979      BY SIMILARITY.
FT   DISULFID    973    988      BY SIMILARITY.
FT   DISULFID    990    999      BY SIMILARITY.
FT   DISULFID   1006   1019      BY SIMILARITY.
FT   DISULFID   1011   1028      BY SIMILARITY.
FT   DISULFID   1030   1039      BY SIMILARITY.
FT   DISULFID   1046   1057      BY SIMILARITY.
FT   DISULFID   1051   1069      BY SIMILARITY.
FT   DISULFID   1071   1080      BY SIMILARITY.
FT   DISULFID   1087   1098      BY SIMILARITY.
FT   DISULFID   1092   1110      BY SIMILARITY.
FT   DISULFID   1112   1121      BY SIMILARITY.
FT   DISULFID   1130   1142      BY SIMILARITY.
FT   DISULFID   1136   1155      BY SIMILARITY.
FT   DISULFID   1157   1166      BY SIMILARITY.
FT   CARBOHYD    711    711      POTENTIAL.
FT   CARBOHYD    960    960      POTENTIAL.
FT   CARBOHYD   1139   1139      POTENTIAL.
FT   CONFLICT     43     43      Q -> R (IN REF. 3).
FT   CONFLICT    298    298      L -> P (IN REF. 3).
FT   CONFLICT    884    884      M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match           5.5%;  Score 622;  DB 1;  Length 1964;
  Best Local Similarity  36.3%;  Pred. No. 2.20e-102;
  Matches   87;  Conservative   55;  Mismatches  87;  Indels  11;  Gaps  10;

Db    692 CISTPCAHGGTCHPQPSG-YNCTCPAGYMGLTCSEEVTACHSGPCLNGGSCSIRP--E-G 747
          |:| || : |||:::|   | |||| |::|  |   : || | || :||:| ::   | |
Qy    918 CLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDG 977

Db    748 YSCTCLPSHTGRHCQTAVDHCVSASCLNGGTCVNKPGTFFCLCATGFQGLHCEEKTNPSC 807
          : | |  :  | :|:  || |    | | :|||:    : |||:  : |  |||| :  |
Qy    978 FWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD-FC 1036

Db    808 A-D-SPCRNKATCQDTPRGARCLCSPGYTGSSCQTLIDLCARKPCPHTARCLQSGPSFQC 865
          | | :||:: : |  ||:| :| |:||| |  |:  :| | : | : |:| ::  :: |
Qy   1037 AQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTC 1096

Db    866 LCLQGWTGALCDFPLSCQMAAMSQGIEISGL-CQNGGLCIDTGSSYFCRCPPGFQGKLCQ 924
          :| :| :| :|:|: :  | :::     : ||||: ||   :  :|:| ||:||  |:
Qy   1097 ICPEGYSGLFCEFS-P-PMV-LPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT  15
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
```

```
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1     90
FT   CHAIN       91   2139       CRUMBS PROTEIN.
FT   DOMAIN      91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085   2111       POTENTIAL.
FT   DOMAIN    2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267    303       EGF-LIKE 1.
FT   DOMAIN     306    343       EGF-LIKE 2.
FT   DOMAIN     348    386       EGF-LIKE 3.
FT   DOMAIN     388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427    463       EGF-LIKE 5.
FT   DOMAIN     464    500       EGF-LIKE 6.
FT   DOMAIN     501    532       EGF-LIKE 7.
FT   DOMAIN     545    581       EGF-LIKE 8.
FT   DOMAIN     582    611       EGF-LIKE 9.
FT   DOMAIN     609    646       EGF-LIKE 10.
FT   DOMAIN     648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840    902       EGF-LIKE 16.
FT   DOMAIN     904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980   1021       EGF-LIKE 19.
FT   DOMAIN    1207   1243       EGF-LIKE 20.
FT   DOMAIN    1481   1517       EGF-LIKE 21.
FT   DOMAIN    1759   1795       EGF-LIKE 22.
FT   DOMAIN    1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874   1915       EGF-LIKE 25.
FT   DOMAIN    1915   1951       EGF-LIKE 26.
FT   DOMAIN    1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030   2070       EGF-LIKE 29.
FT   DISULFID   271    282       BY SIMILARITY.
FT   DISULFID   276    291       BY SIMILARITY.
FT   DISULFID   293    302       BY SIMILARITY.
FT   DISULFID   310    321       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   333    342       BY SIMILARITY.
FT   DISULFID   352    363       BY SIMILARITY.
FT   DISULFID   357    374       BY SIMILARITY.
FT   DISULFID   376    385       BY SIMILARITY.
FT   DISULFID   392    403       BY SIMILARITY.
FT   DISULFID   397    412       BY SIMILARITY.
FT   DISULFID   414    424       BY SIMILARITY.
FT   DISULFID   431    442       BY SIMILARITY.
FT   DISULFID   436    451       BY SIMILARITY.
FT   DISULFID   453    462       BY SIMILARITY.
FT   DISULFID   468    479       BY SIMILARITY.
FT   DISULFID   473    488       BY SIMILARITY.
FT   DISULFID   490    499       BY SIMILARITY.
FT   DISULFID   505    515       BY SIMILARITY.
FT   DISULFID   509    520       BY SIMILARITY.
FT   DISULFID   522    531       BY SIMILARITY.
FT   DISULFID   549    562       BY SIMILARITY.
FT   DISULFID   556    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   586    597       BY SIMILARITY.
FT   DISULFID   591    602       BY SIMILARITY.
FT   DISULFID   604    610       BY SIMILARITY.
FT   DISULFID   613    624       BY SIMILARITY.
FT   DISULFID   618    634       BY SIMILARITY.
FT   DISULFID   636    645       BY SIMILARITY.
FT   DISULFID   652    664       BY SIMILARITY.
FT   DISULFID   659    673       BY SIMILARITY.
FT   DISULFID   675    684       BY SIMILARITY.
FT   DISULFID   691    702       BY SIMILARITY.
FT   DISULFID   696    711       BY SIMILARITY.
FT   DISULFID   713    722       BY SIMILARITY.
FT   DISULFID   729    740       BY SIMILARITY.
FT   DISULFID   734    749       BY SIMILARITY.
FT   DISULFID   751    760       BY SIMILARITY.
FT   DISULFID   767    778       BY SIMILARITY.
FT   DISULFID   772    787       BY SIMILARITY.
FT   DISULFID   789    799       BY SIMILARITY.
FT   DISULFID   806    817       BY SIMILARITY.
FT   DISULFID   811    826       BY SIMILARITY.
FT   DISULFID   828    837       BY SIMILARITY.
FT   DISULFID   844    855       BY SIMILARITY.
FT   DISULFID   849    890       BY SIMILARITY.
FT   DISULFID   892    901       BY SIMILARITY.
FT   DISULFID   908    919       BY SIMILARITY.
FT   DISULFID   913    928       BY SIMILARITY.
FT   DISULFID   930    939       BY SIMILARITY.
FT   DISULFID   946    957       BY SIMILARITY.
FT   DISULFID   952    966       BY SIMILARITY.
FT   DISULFID   968    977       BY SIMILARITY.
FT   DISULFID   984    995       BY SIMILARITY.
FT   DISULFID   989   1009       BY SIMILARITY.
FT   DISULFID  1011   1020       BY SIMILARITY.
FT   DISULFID  1211   1222       BY SIMILARITY.
FT   DISULFID  1216   1231       BY SIMILARITY.
FT   DISULFID  1233   1242       BY SIMILARITY.
FT   DISULFID  1485   1496       BY SIMILARITY.
FT   DISULFID  1490   1505       BY SIMILARITY.
FT   DISULFID  1507   1516       BY SIMILARITY.
FT   DISULFID  1763   1774       BY SIMILARITY.
```

```
FT   DISULFID   1768   1783       BY SIMILARITY.
FT   DISULFID   1785   1794       BY SIMILARITY.
FT   DISULFID   1801   1812       BY SIMILARITY.
FT   DISULFID   1806   1821       BY SIMILARITY.
FT   DISULFID   1823   1832       BY SIMILARITY.
FT   DISULFID   1839   1850       BY SIMILARITY.
FT   DISULFID   1844   1859       BY SIMILARITY.
FT   DISULFID   1861   1870       BY SIMILARITY.
FT   DISULFID   1878   1889       BY SIMILARITY.
FT   DISULFID   1883   1903       BY SIMILARITY.
FT   DISULFID   1905   1914       BY SIMILARITY.
FT   DISULFID   1919   1930       BY SIMILARITY.
FT   DISULFID   1924   1939       BY SIMILARITY.
FT   DISULFID   1941   1950       BY SIMILARITY.
FT   DISULFID   1957   1968       BY SIMILARITY.
FT   DISULFID   1962   1977       BY SIMILARITY.
FT   DISULFID   1979   1988       BY SIMILARITY.
FT   DISULFID   1995   2008       BY SIMILARITY.
FT   DISULFID   2002   2017       BY SIMILARITY.
FT   DISULFID   2019   2028       BY SIMILARITY.
FT   CARBOHYD     37     37       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    238    238       POTENTIAL.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    336    336       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CARBOHYD    550    550       POTENTIAL.
FT   CARBOHYD    565    565       POTENTIAL.
FT   CARBOHYD    736    736       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    860    860       POTENTIAL.
FT   CARBOHYD    884    884       POTENTIAL.
FT   CARBOHYD    976    976       POTENTIAL.
FT   CARBOHYD   1102   1102       POTENTIAL.
FT   CARBOHYD   1114   1114       POTENTIAL.
FT   CARBOHYD   1138   1138       POTENTIAL.
FT   CARBOHYD   1192   1192       POTENTIAL.
FT   CARBOHYD   1245   1245       POTENTIAL.
FT   CARBOHYD   1255   1255       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   CARBOHYD   1363   1363       POTENTIAL.
FT   CARBOHYD   1441   1441       POTENTIAL.
FT   CARBOHYD   1454   1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match           5.4%;  Score 609;  DB 1;  Length 2139;
  Best Local Similarity  44.4%;  Pred. No. 1.67e-99;
  Matches   83;  Conservative   34;  Mismatches  62;  Indels   8;  Gaps   5;

Db    659 C-GNGICKNEKGS-YKCYCTPGFTGVHCDSDVDECLSFPCLNGATCHNKI---NAYECVC 713
          |  :| |:::    |:| |  || |  ||  :  |:| || :|:||| |    ::: |:|
Qy    923 CKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCIC 982

Db    714 QPGYEGENCEVDIDECGSNPCSNGSTCIDRINNFTCNCIPGMRGRICDIDIDDCVGD--P 771
            |:|||||||::|:|  | | | |||:| |||:|| | |   | :|:  :| |  |  |
Qy    983 ADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLNP 1042

Db    772 CLNGGQCIDQLGGFRCDCSGTGYEGENCELNIDECLSNPCTNGAKCLDRVKDYFCDCHNG 831
          | : : ||    ||:|||:  || ||:|::::|:|  | | ||| | | |: | | | :|
Qy   1043 CQHDSKCILTPKGFKCDCT-PGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEG 1101

Db    832 YKGKNCE 838
          | |  ||
Qy   1102 YSGLFCE 1108


Search completed: Fri May 28 08:31:05 1999
Job time : 125 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:31:23 1999; MasPar time 86.17 Seconds
965.924 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-2
Description: (1-1525) from US09191647.pep
Perfect Score: 11299
Sequence: 1 MRGVGWQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 54.443; Variance 108.152; scale 0.503

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	8111	71.8	1523	11	O88280	MEGF5.	0.00e+00
2	7766	68.7	1531	11	O88279	MEGF4.	0.00e+00
3	3678	32.6	739	4	O75094	MEGF5 (FRAGMENT).	0.00e+00
4	1470	13.0	530	5	Q24526	SLIT LOCUS ENCODING A	3.00e-268
5	1107	9.8	601	5	Q20204	F40E10.4 PROTEIN (FRAG	3.88e-192
6	748	6.6	406	5	Q25059	FIBROPELLIN III (FRAGM	3.88e-118
7	723	6.4	529	5	Q25058	FIBROPELLIN IA (FRAGME	4.63e-113
8	704	6.2	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	3.27e-109
9	687	6.1	2531	5	O16004	NOTCH HOMOLOG.	8.95e-106
10	680	6.0	2470	11	O35516	CELL SURFACE PROTEIN.	2.32e-104
11	666	5.9	728	13	Q90656	TRANSMEMBRANE PROTEIN	1.54e-101
12	666	5.9	2352	5	O61240	HRNOTCH PROTEIN.	1.54e-101
13	672	5.9	2447	13	O13149	NOTCH 2 (FRAGMENT).	9.53e-103
14	657	5.8	615	13	O57409	DELTAB.	1.00e-99
15	659	5.8	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	3.96e-100
16	652	5.8	1218	4	O15122	JAGGED1.	1.01e-98
17	650	5.8	1218	4	O14902	TRANSMEMBRANE PROTEIN	2.56e-98
18	650	5.8	1218	4	Q15816	TRANSMEMBRANE PROTEIN	2.56e-98
19	650	5.8	1219	11	Q63722	JAGGED PROTEIN.	2.56e-98
20	650	5.8	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	2.56e-98
21	640	5.7	642	13	P79941	NOTCH LIGAND X-DELTA-2	2.61e-96
22	649	5.7	721	13	Q91902	X-DELTA-1.	4.07e-98
23	648	5.7	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	6.46e-98
24	647	5.7	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	1.03e-97
25	625	5.5	802	13	O57462	DELTAA.	2.66e-93
26	617	5.5	1964	11	O35442	NOTCH4.	1.06e-91
27	613	5.4	717	13	P87357	DELTAD TRANSMEMBRANE P	6.72e-91
28	615	5.4	752	13	O42374	NOTCH RECEPTOR PROTEIN	2.67e-91
29	612	5.4	955	4	Q99466	NOTCH4 (FRAGMENT).	1.06e-90
30	611	5.4	1999	4	Q99940	NOTCH4.	1.69e-90
31	612	5.4	2003	4	O00306	NOTCH4.	1.06e-90
32	595	5.3	1476	13	Q90285	PUTATIVE EXTRACELLULAR	2.64e-87
33	591	5.2	1202	11	P97607	JAGGED2 (FRAGMENT).	1.66e-86
34	591	5.2	1372	5	P91526	SIMILARITY TO MULTIPLE	1.66e-86
35	575	5.1	434	11	O55139	JAGGED2 PROTEIN (FRAGM	2.53e-83
36	554	4.9	832	5	Q99108	NEUROGENIC LOCUS DELTA	3.71e-79
37	553	4.9	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	5.85e-79
38	538	4.8	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	5.40e-76
39	533	4.7	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	5.24e-75
40	507	4.5	473	5	Q25464	ADHESIVE PLAQUE MATRIX	6.79e-70
41	505	4.5	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.68e-69
42	496	4.4	585	11	O35675	M-DELTA-LIKE 3 GENE PR	9.69e-68
43	494	4.4	589	11	O88671	DELTA 3.	2.39e-67
44	492	4.4	592	11	O88516	DELTA-LIKE 3 ALTERNATE	5.87e-67
45	486	4.3	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	8.71e-66

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 71.8%; Score 8111; DB 11; Length 1523;
Best Local Similarity 66.7%; Pred. No. 0.00e+00;
Matches 1011; Conservative 291; Mismatches 203; Indels 10; Gaps 9;

```
Db    16 LALALALASILSGPPAAACPTKCTCSAASVDCHGLGLRAVPRGIPRNAERLDLDRNNITR 75
         |:|:| || ||:  :: |||: |:||:::|||||:||:||| ||||:|||||: |||||
Qy    11 LSLGLVLA-ILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITR 69

Db    76 ITKMDFTGLKNLRVLHLEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKL 135
           ||| ||:||::||||:| :| :| |||||||||:|||||||:| ||::||||| :|:||
Qy    70 ITKTDFAGLRHLRVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKL 129

Db   136 TRLDLSENQIQGIPRKAFRGVTGVKNLQLDNNHISCIEDGAFRALRDLEILTLNNNNISR 195
          |||||||||:|||||||   :|||||| |:|||||||||||||||:|||||||:|
Qy   130 YRLDLSENQIQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITR 189

Db   196 ILVTSFNHMPKIRTLRLHSNHLYCDCHLAWLSDWLRQRRTIGQFTLCMAPVHLRGFSVAD 255
          : |:|||||||:||:|||||:|||||||||||||| |  :| :| ||:| |||| :||:
```

```
Qy   190 LSVASFNHMPKLRTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAE 249

Db   256 VQKKEYVCPGPH-S-EA-PACNANSLSCPSACSCSNNIVDCRGKGLTEIPANLPEGIVEI 312
Qy   250 VQKREFVCSDEEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEI 309

Db   313 RLEQNSIKSIPAGAFIQYKKLKRIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIPK 372
Qy   310 RLEQNTIKVIPPGAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPK 369

Db   373 GLFDGLVSLQLLLLNANKINCLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTL 432
Qy   370 SLFEGLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTM 429

Db   433 HLAQNPFVCDCHLKWLADYLQDNPIETSGARCSSPRRLANKRISQIKSKKFRCSGSEDYR 492
Qy   430 HLAQNPFICDCHLKWLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYR 489

Db   493 NRFSSECFMDLVCPEKCRCEGTIVDCSNQKLSRIPSHLPEYTTDLRLNDNDIAVLEATGI 552
Qy   490 SKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGI 549

Db   553 FKKLPNLRKINLSNNRIKEVREGAFDGAAGVQELMLTGNQLETMHGRMFRGLSGLKTLML 612
Qy   550 FKKLPQLRKINFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLML 609

Db   613 RSNLISCVNNDTFAGLSSVRLLSLYDNRITTISPGAFTTLVSLSTINLLSNPFNCNCHMA 672
Qy   610 RSNRITCVGNDSFIGLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLA 669

Db   673 WLGRWLRKRRIVSGNPRCQKPFFLKEIPIQDVAIQDFTCE-GNEENSCQLSPRCPEQCTC 731
Qy   670 WLGEWLRKKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTC 729

Db   732 VETVVRCSNRGLHTLPKGMPKDVTELYLEGNHLTAVPKELSTFRQLTLIDLSNNSISMLT 791
Qy   730 LDTVVRCSNKGLKVLPKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLS 789

Db   792 NHTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSLSH 851
Qy   790 NQSFSNMTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSH 849

Db   852 LALGINPLHCDCSLRWLSEWIKAGYKEPGIARCSSPESMADRLLLTTPTHRFQCKGPVDI 911
Qy   850 LAIGANPLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDV 909

Db   912 NIVAKCNACLSSPCKNNGTCSQDPVEQYRCTCPYSYKGKDCTVPINTCVQNPCQHGGTCH 971
Qy   910 NILAKCNPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCH 969

Db   972 LSESHRDGFSCSCPLGFEGQRCEINPDDCEDNDCENSATCVDGINNYACVCPPNYTGELC 1031
Qy   970 LKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELC 1029

Db  1032 DEVIDYCVPEMNLCQHEAKCISLDKGFRCECVPGYSGKLCETDNDDCVAHKCRHGAQCVD 1091
Qy  1030 EEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTD 1089

Db  1092 AVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQQEPTCRCPPGFAG 1151
Qy  1090 AVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQG 1149

Db  1152 PRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYKGDNDPLALELYQG 1211
Qy  1150 EKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRG 1209

Db  1212 HVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVMLNQTLNLVVDKGAPKSLGKLQKQPAV 1271
Qy  1210 RVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTL 1269

Db  1272 GINSPLYLGGIPTSTGLSALRQGADRPLGGFHGCIHEVRINNELQDFKALPPQSLGVSPG 1331
Qy  1270 NFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQT-GILPG 1328

Db  1332 CKSC--TVCRHGLCRSVEKDSVVCECHPGWTGPLCDQEAQDPCLGHSCSHGTCVATGN-S 1388
Qy  1329 CEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFS 1388

Db  1389 YVCKCAEGYEGPLCDQKNDSANACSAFKCHHGQCHISDRGEPYCLCQPGFSGNHCEQENP 1448
Qy  1389 YSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYCECSSGYTGDSCDREIS 1448

Db  1449 CLGEIVREAIRRQKDYASCATASKVPIMVCRGGC-GSQCCQPIRSKRRKYVFQCTDGSSF 1507
Qy  1449 CRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSF 1508

Db  1508 VEEVERHLECGCREC 1522
Qy  1509 VDEVEKVVKCGCTRC 1523


RESULT   2
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match          68.7%;  Score 7766;  DB 11;  Length 1531;
  Best Local Similarity  65.3%;  Pred. No. 0.00e+00;
  Matches  984;  Conservative  284;  Mismatches 229;  Indels  11;  Gaps   7;

Db    28 RLGATACPALCTCTGTTVDCHGTGLQAIPKNIPRNTERLELNGNNITRIHKNDFAGLKQL 87
Qy    22 KVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHL 81

Db    88 RVLQLMENQIGAVERGAFDDMKELERLRLNRNQLQVLPELLFQNNQALSRLDLSENSLQA 147
Qy    82 RVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQA 141

Db   148 VPRKAFRGATDLKNLQLDKNQISCIEEGAFRALRGLEVLTLNNNNITTIPVSSFNHMPKL 207
Qy   142 IPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKL 201

Db   208 RTFRLHSNHLFCDCHLAWLSQWLRQRPTIGLFTQCSGPASLRGLNVAEVQKSEFSCSGQG 267
Qy   202 RTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEE 261

Db   268 EAAQV-PACTLSSGSCPAMCSCSNGIVDCRGKGLTAIPANLPETMTEIRLELNGIKSIPP 326
Qy   262 EGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPP 321

Db   327 GAFSPYRKLRRIDLSNNQIAEIAPDAFQGLRSLNSLVLYGNKITDLPRGVFGGLYTLQLL 386
Qy   322 GAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLL 381

Db   387 LLNANKINCIRPDAFQDLQNLSLLSLYDNKIQSLAKGTFTSLRAIQTLHLAQNPFICDCN 446
```

```
Qy   382 LLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCH 441

Db   447 LKWLADFLRTNPIETTGARCASPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYHLNSE 506

Qy   442 LKWLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDY--RSK----LSGD 495

Db   507 CTSDVACPHKCRCEASVVECSGLKLSKIPERIPQSTTELRLNNNEISILEATGLFKKLSH 566

Qy   496 CFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQ 555

Db   567 LKKINLSNNKVSEIEDGTFEGATSVSELHLTANQLESVRSGMFRGLDGLRTLMLRNNRIS 626

Qy   556 LRKINFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRIT 615

Db   627 CIHNDSFTGLRNVRLLSLYDNHITTISPGAFDTLQALSTLNLLANPFNCNCQLAWLGDWL 686

Qy   616 CVGNDSFIGLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWL 675

Db   687 RKRKIVTGNPRCQNPDFLRQIPLQDVAFPDFRCEEGQEEVGCLPRPQCPQECACLDTVVR 746

Qy   676 RKKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVR 735

Db   747 CSNKHLQALPKGIPKNVTELYLDGNQFTLVPGQLSTFKYLQLVDLSNNKISSLSNSSFTN 806

Qy   736 CSNKGLKVLPKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQSFSN 795

Db   807 MSQLTTLILSYNALQCIPPLAFQGLRSLRLLSLHGNDVSTLQEGIFADVTSLSHLAIGAN 866

Qy   796 MTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGAN 855

Db   867 PLYCDCHLRWLSSWVKTGYKEPGIARCAGPPEMEGKLLLTTPAKKFECQGPPSLAVQAKC 926

Qy   856 PLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKC 915

Db   927 DPCLSSPCQNQGTCHNDPLEVYRCTCPSGYKGRNCEVSLDSCSSNPCGNGGTCHAQEGED 986

Qy   916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db   987 AGFTCSCPSGFEGLTCGMNTDDCVKHDCVNGGVCVDGIGNYTCQCPLQYTGRACEQLVDF 1046

Qy   976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db  1047 CSPDLNPCQHEAQCVGTPEGPRCECVPGYTGDNCSKNQDDCKDHQCQNGAQCVDEINSYA 1106

Qy  1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db  1107 CLCAEGYSGQLCEI-PP-A-PRNS-CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 1162

Qy  1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKL 1155

Db  1163 LSVNFVDRDTYLQFTDLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELYQGHVRVSY 1222

Qy  1156 VSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRASY 1215

Db  1223 DPGSYPSSAIYSAETINDGQFHTVELVTFDQMVNLSIDGGSPMTMDNFGKHYTLNSEAPL 1282

Qy  1216 DTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPL 1275

Db  1283 YVGGMPVDVNSAAFRLWQILNGTSFHGCIRNLYINNELQDFTKTQMKPGVVPGCEPCRKL 1342

Qy  1276 YVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPGCEPCHKK 1335

Db  1343 YCLHGICQPNATPGPVCHCEAGWGGLHCDQPVDGPCHGHKCVHGKCVPLDALAYSCQCQD 1402

Qy  1336 VCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLE 1395

Db  1403 GYSGALCNQVGAVAEPCGGLQCLHGHCQASATRGAHCVCSPGFSGELCEQESECRGDPVR 1462

Qy  1396 GHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYCECSSGYTGDSCDREISCRGERIR 1455

Db  1463 DFHRVQRGYAICQTTRPLSWVECRGACPGQGCCQGLRLKRRKLTFECSDGTSFAEEVEKP 1522

Qy  1456 DYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKV 1515

Db  1523 TKCGCAPC 1530

Qy  1516 VKCGCTRC 1523
```

RESULT 3

ID O75094 PRELIMINARY; PRT; 739 AA.
AC O75094;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5 (FRAGMENT).
GN MEGF5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011538; D1033429; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 739 AA; 80364 MW; DC6BCB63 CRC32;

Query Match 32.6%; Score 3678; DB 4; Length 739;
Best Local Similarity 59.8%; Pred. No. 0.00e+00;
Matches 444; Conservative 156; Mismatches 137; Indels 5; Gaps 4;

```
Db     1 NSISMLTNYTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFN 60

Qy   783 NRISTLSNQSFSNMTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFN 842

Db    61 DLTSLSHLALGTNPLHCDCSLRWLSEWVKAGYKEPGIARCSSPEPMADRLLLTTPTHRFQ 120

Qy   843 DLSALSHLAIGANPLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFT 902

Db   121 CKGPVDINIVAKCNACLSSPCKNNGTCTQDPVELYRCACPYSYKGKDCTVPINTCIQNPC 180

Qy   903 CQGPVDVNILAKCNPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPC 962

Db   181 QHGGTCHLSDSHKDGFSCSCPLGFEGQRCEINPDDCEDNDCENNATCVDGINNYVCICPP 240

Qy   963 KHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPP 1022

Db   241 NYTGELCDEVIDHCVPELNLCQHEAKCIPLDKGFSCECVPGYSGKLCETDNDDCVAHKCR 300

Qy  1023 EYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCK 1082

Db   301 HGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQQEPTCR 360

Qy  1083 NGAHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQ 1142

Db   361 CPPGFAGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYKGDNDPL 420

Qy  1143 CLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHI 1202

Db   421 ALELYQGHVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVTLNQTLNLVVDKGTPKSLGK 480

Qy  1203 AVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITN 1262

Db   481 LQKQPAVGINSPLYLGGIPTSTGLSALRQGTDRPLGGFHGCIHEVRINNELQDFKALPPQ 540

Qy  1263 LSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQ 1322
```

```
Db    541 SLGVSPGCKSC--TVCKHGLCRSVEKDSVVCECRPGWTGPLCDQEARDPCLGHRCHHGKC 598
          : |: ||| :|   || || |::    :  |||: || |||||| : |||||::| || |
Qy   1323 T-GILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTC 1381

Db    599 VATGT-SYMCKCAEGYGGDLCDNKNDSANACSAFKCHHGQCHISDQGEPYCLCQPGFSGE 657
          ::  : || ||| || || |||: :|  |:| |:|| || |::|  |:||| | :|::|:
Qy   1382 LPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYCECSSGYTGD 1441

Db    658 HCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQ 716
           |::| :| |: :|:  ::| |||:| |: ||: :|||||| | ||| | |||||| |:
Qy   1442 SCDREISCRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFE 1501

Db    717 CTDGSSFVEEVERHLECGCLAC 738
          ||||||||:|||: : |||  |
Qy   1502 CTDGSSFVDEVEKVVKCGCTRC 1523


RESULT    4
ID   Q24526      PRELIMINARY;      PRT;   530 AA.
AC   Q24526;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SLIT LOCUS ENCODING A PROTEIN ASSOCIATED WITH NEURAL DEVELOPMENT WITH
DE   52D EGF HOMOLOGOUS DOMAINS (FRAGMENT).
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 89077533.
RA   ROTHBERG J.M., HARTLEY D.A., WALTHER Z., ARTAVANIS-TSAKONAS S.;
RT   "slit: an EGF-homologous locus of D. melanogaster involved in the
RT   development of the embryonic central nervous system.";
RL   CELL 55:1047-1059(1988).
DR   EMBL; M23543; G514357; -.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
KW   NEUROGENESIS; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     530    530
SQ   SEQUENCE   530 AA;  59457 MW;  10E5764D CRC32;

  Query Match           13.0%;  Score 1470;  DB 5;  Length 530;
  Best Local Similarity  39.1%;  Pred. No. 3.00e-268;
  Matches   206;  Conservative  132;  Mismatches 168;  Indels  21;  Gaps  17;

Db      1 MKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHG 60
          | |||:|:|||  | |:| |  :|||||:|:  || |:: | : |   |:| |  |: |
Qy    888 MADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKG 947

Db     61 KHCEFMIDACYGNPCRNNATCTVLE--EGRFSCQCAPGYTGARCETNIDDCLGEIKCQNN 118
            |:  | || :|||:: :|| : |  |  | | || |: |  || |:|||  :  |:||
Qy    948 QDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDC-EDNDCENN 1006

Db    119 ATCIDGVESYKCECQPGFSGEFCDTKIQFCSPEFNPCANGAKCMDHFTHYSCDCQAGFHG 178
          :|:||:::| | | | ::||:|: |::||: ::||| : :||:     : ||| :|: |
Qy   1007 STCVDGINNYTCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVG 1066

Db    179 TNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNH 238
           :|  ::||||:::| ||: | |::| | | ||: |:| :||  :  :|: |:||||:|
Qy   1067 EHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEGYSGLFCE-FSP-PMVLPRTSPCDNF 1124

Db    239 ECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTRPEANVTI 298
          :| :|   |   : :: :|:| ||| |  || | |::|::::|:::: : : ||::|:|:
Qy   1125 DCQNGA--QCIVRINEPICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITL 1182

Db    299 VFSSGQN-GILMYDGQDAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVELL 357
          ::: :: |||:| |:  |:||||: ||:| ||| |:|| |::|| | : ||::| ||||
Qy   1183 QIATDEDSGILLYKGDKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1242

Db    358 AIKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFLGGLPVDPAQQAYKNWQIRNLTSFKG 417
          |:  :::| || | :: | | :    |:: :|:::||:|  :   : :    :| ||| |
Qy   1243 ALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHG 1302

Db    418 CMKEVWINHKLVDFGNAQRQQKITPGCAL-LEGE-QQEE-EDDEQD-FM-D--ET--PHI 468
          |:::: ||  | || :  |  | |||         :   :   |  |  :  |     :
Qy   1303 CIRNLYINSELQDFQKVPMQTGILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPL 1362

Db    469 KEEPV-DPCLENKCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQGE 514
          ::   |||| ||| :|: |:|  ||  :| |||  |: |  ||: |
Qy   1363 CDQRTNDPCLGNKCVHGT-CLPI-NAF-SYSCKCLEGHGGVLCDEEE 1406


RESULT    5
ID   Q20204      PRELIMINARY;      PRT;   601 AA.
AC   Q20204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F40E10.4 PROTEIN (FRAGMENT).
GN   F40E10.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMYE R.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z69792; E1346469; -.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match           9.8%;  Score 1107;  DB 5;  Length 601;
  Best Local Similarity  34.7%;  Pred. No. 3.88e-192;
  Matches   182;  Conservative  124;  Mismatches 186;  Indels  32;  Gaps  16;

Db      1 IKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAIC 60
          :|| : |:|||||  |  ::: |||: :  |||:: |  :: :||: ||::|||:: |
Qy    870 VKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNPCKNDGTC 929

Db     61 ETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATC--KVAQAGRFNCYCNKGFEGD 118
          :: :  | | |  || |  |:  | || ::|| : :||  | ::   | | |  ||||:
Qy    930 NSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWCICADGFEGE 989

Db    119 YCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLE 178
           || |:||| :: ||| : |||            :  ||:| | || || |  ||:||:
Qy    990 NCEVNVDDCEDNDCENNSTCVD----------G--INNYTCLCPPEYTGELCEEKLD 1034

Db    179 YCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSY 238
          :|: ||||::::|||   :: | |:||: |::|: ::||| :  | ||: | |:: :|
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094
```

```
Db    239 DCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQG-ECVASQNSSDFTCKCHEGFSGP 297
          |:|  ||:| :||::| | :   :| :|:   | | :|:   | : : | |  |: |
Qy   1095 TCICPEGYSGLFCEFSPPMVL--PRTSPCDNFDCQNGAQCIVRINEP-I-CQCLPGYQGE 1150

Db    298 SCDRQMSVGFKNPGAYLAL-DP-LASDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGR 355
          |:: :|| | |  :|| :  : : ::  ||: : |    ||||| ||   :: ||| ||
Qy   1151 KCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGR 1210

Db    356 VKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVENSGKSDQLI 415
          |:  |  |: ||| :||   :||| | : : : ::   | :|     |: | :|   |
Qy   1211 VRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLN 1270

Db    416 TKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQS 475
            :  ||:||:|   :  : |:    |: |: ||| :: ||  : :||: :::|  : :
Qy   1271 FDS-P-LYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYIN--S-ELQD-FQKVPMQTG 1324

Db    476 CSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDEK 519
            :  : |    |::| | |  : | |: |:|:    |  ||::
Qy   1325 ILPGCEPCHKKVCAHGTC-QPS-SQAGFTCECQEGWMGPLCDQR 1366


RESULT    6
ID   Q25059      PRELIMINARY;      PRT;   406 AA.
AC   Q25059;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN III (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33862; G499688; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 7.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   406 AA;  43475 MW;  45E6EE2C CRC32;

  Query Match           6.6%;  Score 748;  DB 5;  Length 406;
  Best Local Similarity  42.3%;  Pred. No. 3.88e-118;
  Matches  101;  Conservative   47;  Mismatches  76;  Indels  15;  Gaps  11;

Db     14 DDCNPNPCQNGAACI-DQVNDYECICPPGFTGDNCETDIDVCASAPCRNGGAC-V-DGV- 69
          : | :||| | ::|  | |: | | | || || |::|:  |  | | ||::||:| : :|
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db     70 NGYTCNCIPGFDGDNCENNINECASNPCQNGGVCIDGVNGFVCTCQPGYTGTLCETDIDE 129
          :|: | |  ||:|:||| |:::|  | |:| : |:||:| : | | | ||| |||  :|
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    130 CA-S-NPCQNGGVCTDLVNM-YTCDCLAGFTGSNCETNINECASNPCLNGGACVDGVNGY 186
          ||  ||||: : |  |  : : ||| :|: | :|: :::|  | | ||: | |:||||
Qy   1036 CAQDLNPCQHDSKCI-LTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    187 VCQCLPNYTGTHCEIS----LD-V--CQSMPCQNGATCTNVGGDYSCECPPGYTGINCE 238
          | |   |:|  ||:|    |     |::  ||||| |     :  |:| ||| | :||
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT    7
ID   Q25058      PRELIMINARY;      PRT;   529 AA.
AC   Q25058;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN IA (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33861; G499686; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 10.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   529 AA;  55543 MW;  6385F322 CRC32;

  Query Match           6.4%;  Score 723;  DB 5;  Length 529;
  Best Local Similarity  40.2%;  Pred. No. 4.63e-113;
  Matches   96;  Conservative   49;  Mismatches  79;  Indels  15;  Gaps   9;

Db    137 DECASSPCLNGGQCI-NRINSYECVCAAGFNGVNCQTNIDECASDPCENGGIC-I-AGV- 192
          : | |:|| | | |  : :: | | |: ||:| :|:  |  | |:|| :|| | :  |
Qy    916 NPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEE 975

Db    193 NGYTCNCASGYTGTNCETEIDECASMPCLNGGQCIEMVNGYTCQCAAGFTGVLCETDIDE 252
          :|: | || |: | ||| ::|:|    | | : |:: :| ||| |:: :|| |||  :|
Qy    976 DGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDF 1035

Db    253 CASD--PCQNGGVCTDTVNGYICSCVQGFTGSDCETNINECASGPCQNGGTCVDGVNGFV 310
          || |  |||: : |  | :|: | |  |: |  |: :::|    | ||: | |:|||:
Qy   1036 CAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYT 1095

Db    311 CQCPPNYTGTYCEIS----L---DACSSMPCQNGATC-VNVGANYICECPPGFAGQNCE 361
          | ||  |:| :||:|    |    :| :  ||||| | | :    ||:| ||: |::||
Qy   1096 CICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEP-ICQCLPGYQGEKCE 1153


RESULT    8
ID   Q25253      PRELIMINARY;      PRT;  2653 AA.
AC   Q25253;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH HOMOLOG SCALLOPED WINGS (SCL).
GN   SCL.
OS   LUCILIA CUPRINA (GREENBOTTLE FLY) (AUSTRALIAN SHEEP BLOWFLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; OESTROIDEA; CALLIPHORIDAE;
OC   LUCILIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SS SEEKING;
RX   MEDLINE; 96400928.
RA   DAVIES A.G., GAME A.Y., CHEN Z., WILLIAMS T.J., GOODALL S., YEN J.L.,
RA   MCKENZIE J.A., BATTERHAM P.;
RT   "Scalloped wings is the Lucilia cuprina Notch homologue and a
RT   candidate for the modifier of fitness and asymmetry of diazinon
RT   resistance.";
RL   GENETICS 143:1321-1337(1996).
RN   [2]
RP   SEQUENCE OF 39-265 FROM N.A.
RC   STRAIN=SS SEEKING;
RA   CHEN Z., NEWSOME T., MCKENZIE J.A., BATTERHAM P.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE OF 39-265 FROM N.A.
RC   STRAIN=SS SEEKING;
RA   CHEN Z., MCKENZIE J.A., BATTERHAM P.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U58977; G1389670; -.
DR   EMBL; AF032672; G2654074; -.
DR   EMBL; AF032670; G2654074; JOINED.
DR   EMBL; AF032671; G2654074; JOINED.
DR   EMBL; AF032673; G2654075; -.
```

```
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2653 AA;  285928 MW;  8F35FD2D CRC32;

  Query Match           6.2%;  Score 704;  DB 5;  Length 2653;
  Best Local Similarity  28.6%;  Pred. No. 3.27e-109;
  Matches  159;  Conservative  131;  Mismatches 215;  Indels  51;  Gaps  45;

Db    392 DACTSNPCHADAICDTSPINGSYTCPCATGYKGVDCSEDIDECDQGSPCEHNGVC-VN-T 449
          ::| |||| |: |:: |::  | | |: |:|| ||   |  |  ::|| | | | ::
Qy    916 NPCLSNPCKNDGTCNSDPVD-FYRCTCPYGFKGQDCDVPIHACI-SNPCKHGGTCHLKEG 973

Db    450 P-GSFRCNCSQGFTGPRCETNINECESHPCQNEGSCLDDPGTFRCVCMPGFTGTQCEIDI 508
            :|:| |::|| |  || |:::|| : |:|:::|:|    : |:| | :|| ||  :
Qy    974 EEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKL 1033

      509 NEC-QS-NPCLNGGICNDMINGFKCSCALGFTGSRCQINIDDCQSQPCRNNGICRDSIAG 566
          : | |  ||| : : |    :|||| |: |: | :|:|::|||   |:| : | |:: |
Qy   1034 DFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNG 1093

Db    567 YTCQCPPGYTGLSCEIN--INDCNSNPCHRGKCIDGDNRFTCVCDPGFTGYLCQTQINEC 624
          ||| || ||:|| ||:: :    ::||    | :|   :  : :| :   |   | : |
Qy   1094 YTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEP-ICQCLPGYQGEKC 1152

Db    625 ESNPCQYGGHCVDRVGSYMCHCLAGTSGKDCEINVN-ECHSNPCNNGATCIDGINKYTCQ 683
          | :    : : :::  ||: : ::::  :  : |:  :  ::  ::|     | :|
Qy   1153 E-KLV--SVNFINK-ESYL-Q-IPSAKVRP-QTNITLQIATDE-DSGILLYKG-DKDHIA 1203

Db    684 CVPGFTGVHCEININECASNPCANNGVCMDLVNGYKCECPRGFYDPRCLSDVDECASNPC 743
          |  :. | :   :  : :|:| |:    :: :|  :      :   : ||   : ::||
Qy   1204 -VELYRG-RVRAS-YDTGSHP-ASAIYSVETINDGNFHIVELLALDQSLSLSVD-GGNPK 1258

Db    744 INGGRCEDGINEFICHCPPGY-GGKRCENDIDECSSNPCQHG-GF--CVDELNAFKCQCM 799
          |      ::  :|    |  | ||:   :::      | |:| :|  |: :|  :: : :
Qy   1259 IITNLSKQSTLNF--DSPL-YVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLY-INSE-L 1313

Db    800 PGYTGLKCETNI--D-D-CINNPCANGGTCIDKVN-GYKCVCKVPYTGQDCESKL-DPCA 853
          :  :  :| |    : | :: ||:| ||      |: | |     |  |: :  |||
Qy   1314 QDFQKVPMQTGILPGCEPCHKKVCAHG-TCQPSSQAGFTCECQEGWMGPLCDQRTNDPCL 1372

Db    854 TNRCRNDAKCTPSPNFLDFSCTCKLGYTGRYCDEDID--E-CKLSTPCRNGATCH-NVPG 909
          |:| : : | |   |  :|| |  |  |  |||: |  : |  :  |::|  |: :  |
Qy   1373 GNKCVHGT-CLPINAF-SYSCKCLEGHGGVLCDEEEDLFNPCQ-AIKCKHG-KCRLSGLG 1428

      910 -SYRCICAKGYEGHDC 924
          :| | |: || |  |
Qy   1429 QPY-CECSSGYTGDSC 1443


RESULT   9
ID   O16004      PRELIMINARY;      PRT;  2531 AA.
AC   O16004;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH HOMOLOG.
OS   LYTECHINUS VARIEGATUS (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; TEMNOPLEUROIDA; TOXOPNEUSTIDAE; LYTECHINUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97454256.
RA   SHERWOOD D.R., MCCLAY D.R.;
RT   "Identification and localization of a sea urchin Notch homologue:
RT   insights into vegetal plate regionalization and Notch receptor
RT   regulation.";
RL   DEVELOPMENT 124:3363-3374(1997).
DR   EMBL; AF000634; G2570351; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 21.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2531 AA;  273982 MW;  BB9C6F3D CRC32;

  Query Match           6.1%;  Score 687;  DB 5;  Length 2531;
  Best Local Similarity  40.3%;  Pred. No. 8.95e-106;
  Matches   96;  Conservative   54;  Mismatches  73;  Indels  15;  Gaps  11;

Db    631 CSSNPCVNDGTC-VDGINEYTCMCHEGYRGLNCEEDIDDCESRPCHNGGTC-V-D-EVNG 686
          | |||| |||||  | :: | | |  |::| :|:  |  | | || :|||| : : | :|
Qy    918 CLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDG 977

Db    687 YHCLCPIGYHDPFCMSNINECSSNPCVNGGSCHDGVNEYSCECMAGYTGTRCTDDFDECS 746
          : |:|: |:    |  |:::|  | | | ::| ||:|:|:| | : |||  | : :| |:
Qy    978 FWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCA 1037

Db    747 -S-NPCQHGGTCDNRHAFYNCTCQAGYTGLNCEVNIDDCVDEPCLNGGICIDEVNSFQCV 804
           |||| : |      ::| | :|| | :|::::||| |: | ||: | | ||:: |:
Qy   1038 QDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCI 1097

Db    805 CPQTFVGLLCE-TE-----R-SPCEDNQCQNGATCVYSEDYAGYSCRCTSGFQGNFCD 855
          ||: : ||:|| :      | |||:: :|||| |:   :     |:| :|:||: |:
Qy   1098 CPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRIN-EPI-CQCLPGYQGEKCE 1153


RESULT   10
ID   O35516      PRELIMINARY;      PRT;  2470 AA.
AC   O35516;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CELL SURFACE PROTEIN.
GN   NOTCH2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RA   HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL   SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D32210; D1022953; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match           6.0%;  Score 680;  DB 11;  Length 2470;
  Best Local Similarity  29.5%;  Pred. No. 2.32e-104;
  Matches  162;  Conservative  116;  Mismatches 220;  Indels  51;  Gaps  40;

Db    184 NECDIPGRCQHGGTCLNLPGS-YRCQCGQGFTGQHCDSP-YV-RGLPCVNGGTCR-QTGD 239
          | | ::  | : ||| : |   ||| |  || || || |     : || :||||:   |:
Qy    916 NPC-LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGE 974

Db    240 FT-LECNCLPGFEGSTCERNIDDCPNHKCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVD 298
```

```
          : | |  ||||  || |:||| :: |:| : ||||:| | | |||: ||::| | :|
Qy    975 EDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD 1034

Db    299 ECLLQPNACQNGGTCTNRNGGYGCVCVNGWSGDDCSENIDDCAYASCTPGSTCIDRVASF 358
          |  : |:||: : |    |: | |  |  |: |  ::|||    |  |: | | | ::
Qy   1035 FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    359 SCLCPEGKAGLLCH------L--DDACISNPCHKGALCDTNPLNGQYICTCPQGYKGADC 410
          :|:|||| :||:|      |   :| :  |::|| |   :|  || |  || |  !
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQC-IVRINEP-ICQCLPGYQGEKC 1152

Db    411 TEDVDECAMANSNPC-E-HAGKC-VNTDGAFHCEC-LK-G---YAGPRCEMDINECHSDP 462
          | :     | ::  :  ::|   |: :::      |  | | :  : ::  ::
Qy   1153 -EKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRV 1211

Db    463 CQN-DATCLDKIGGFTCLCMP-G-FKGVHCELEVNECQSNPC-VNNGQCVDKVNRFQCLC 518
          : |:     : ::   :  | |  |   | :::  | :   |  : ::::  |
Qy   1212 RASYDTGSHPASAIYSVETINDGNFHIVE-LLALDQSLSLSVDGGNPKIITNLSKQSTLN 1270

Db    519 --PPGFTGPVC-QIDIDDCSSTPCLNGAKCIDH-PNGY-ECQCATGFTGILCDENI-DNC 572
          :| : | :   ::    :|  ||:       | | : :   |  :  :  |   |
Qy   1271 FDSPLIVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSE-LQDFQKVPMQTGILPGC 1329

Db    573 DP-DP--CHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 626
          :|    | || || : :: :|| |  | ||::| :: :: | :: |:: | |: :   :
Qy   1330 EPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 1388

Db    627 YQCNCQPGTSGLNC--EIN-FDDCASNPCMHGVC-VDGINR-YSCVCSPGFTGQRCNIDI 681
          | |:|  | :|: |  | : |: | :  |:|| | : |: : | | ||:|:||: |: :|
Qy   1389 YSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPY-CECSSGYTGDSCDREI 1447

Db    682 DECASNPCR 690
          | ::  |
Qy   1448 S-CRGERIR 1455


RESULT  11
ID   Q90656      PRELIMINARY;      PRT;   728 AA.
AC   Q90656;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE PROTEIN C-DELTA-1.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 95319507.
RA   HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT   "Expression of a Delta homologue in prospective neurons in the
RT   chick.";
RL   NATURE 375:787-790(1995).
DR   EMBL; U26590; G882412; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   728 AA;  79861 MW;  7439F575 CRC32;

  Query Match           5.9%;  Score 666;  DB 13;  Length 728;
  Best Local Similarity  39.5%;  Pred. No. 1.54e-101;
  Matches   94;  Conservative   50;  Mismatches  81;  Indels  13;  Gaps   9;

Db    303 KPCKNGATCTNTGQGSYTCSCRPGYTGSSCEIEINECDANPCKNGGSC--TD-LENSYSC 359
          :|||| :|| :    | |:|  |: |  |:: |: | :||||:||:|   :  |::: |
Qy    921 NPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFWC 980

Db    360 TCPPGFYGKNCELSAMTCADGPCFNGGRCTDNPDGGYSCRCPLGYSGFNCEKKIDYCS-S 418
          |: || | |||::  | |  | | : | |  :  |:| ||  |:|  || |:|:|:|:
Qy    981 ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDFCAQD 1039

Db    419 -SPCANGAQCVDLGNSYICQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCP 477
           :|| : : |:   ::: |:| :|: | ||| : |||    | ||: | |:|| |:| ||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    478 PGYNGKNC--STP-V-SR---CEHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
          ||:|  |  | | | :|   |::  |:||| |  | |  :|:|  || | :|: |::
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEKLVS 1157


RESULT  12
ID   O61240      PRELIMINARY;      PRT;  2352 AA.
AC   O61240;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HRNOTCH PROTEIN.
GN   HRNOTCH.
OS   HALOCYNTHIA RORETZI (SEA SQUIRT).
OC   EUKARYOTA; METAZOA; CHORDATA; UROCHORDATA; ASCIDIACEA; STOLIDOBRANCHIA;
OC   PYURIDAE; HALOCYNTHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HORI S., SAITOH T., MATSUMOTO M., MAKABE K.W., NISHIDA H.;
RL   DEV. GENES EVOL. 207:371-380(1997).
DR   EMBL; AB001327; D1026501; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS01186; EGF_2; 22.
DR   PROSITE; PS01187; EGF_CA; 18.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2352 AA;  252623 MW;  816976D4 CRC32;

  Query Match           5.9%;  Score 666;  DB 5;  Length 2352;
  Best Local Similarity  26.7%;  Pred. No. 1.54e-101;
  Matches  147;  Conservative  131;  Mismatches 226;  Indels  47;  Gaps  38;

Db    756 SPCVPNPCENGATCQ-ESADYLAYVCQCPEGFRGPTCATDINECVNSPCKNGGGC--TN- 811
          :||::||| | :||  :: |:  | | || ||:|  |   |: |:::|||:|| |   :
Qy    916 NPCLSNPCKNDGTCNSDPVDF--YRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEG 973

Db    812 LVPGYQCTCSQGFTGKDCDTDIDDCSSNPCLNGGQCLDDVGSYKCLCLPGFEGNNCQEEV 871
            |: | |::|| | :|: ::|||  | | | : |:| : :| ||| | : |: |:| :
Qy    974 EEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKL 1033

Db    872 NECA-SF-PCKNGGICTDYVNSYVCTCLSGFYSLDCEKNIEDCSSSSCMNGGTCVDGINS 929
          : ||  : || : : |    ::: | | :|: :  |: :::||  : |:||: | |::|:
Qy   1034 DFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNG 1093

Db    930 YSCSCTANFTGDKCQ-NA--V---NN-CASLQCQNGGTCYYDSGDPKCACVHGYTGTHCE 982
          |:| |   ::|  |: ::  |    : | :::|||| : |     :| | |: || |  ||
Qy   1094 YTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153

Db    983 SLQNLC-TGPNIC-KNGGSCVQTSNTVSCNCLGGYEGTDCAVPQVSCT-VGASLLGIAVS 1039
          | ::    :      :: |:   ::   ::  | :  :       ::  |    |
Qy   1154 KLVSVNFINKESYLQIPSAKVRPQTNITLQ-IATDEDSGILLYKGDKDHIAVELYRGRVR 1212

Db   1040 DLCLNGGTCHDTSTAHECSCVAGFTGSYCDI-DIDECASVPCKNGATCNDLINSYSCICA 1098
           : |: | :|:      :        ::  :|:  |::  :|:  : :| : |   :
Qy   1213 A-SYDTGS-HPASAIYSVETINDGNFHIVELLALDQSLSLSV-DGGNPK-IITNLSKQST 1268

Db   1099 LGYEGATC---LTDKDECAS---SPCKNGGTCIDRINSFYCSC-LAGTEGV-L-CEI-NE 1148
          | ::::    :  | : ||   :|  || :    | ::| :  |   : | :   |
Qy   1269 LNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTGILPG 1328

Db   1149 -DECEINICLNGGVCI-DGIGGFSCQCPSGYEGRRCQGDVNE-CLSNPCSSPGSLACIQG 1205
           : |  ::| :|  |   : :||:|:|  |  |  |:   |: ||:| |     |: | :
Qy   1329 CEPCHKKVCAHG-TCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLP-INA 1386

Db   1206 SNSYQCVCDADYTGSEC-QIR-I-GSCDINPCLNDGICTDNSQDITGYKCQCTWGYYGKK 1262
           || | |    |  | :  :  :|:  | : | |  ::    | |:|: || |
Qy   1387 F-SYSCKCLEGHGGVLCDEEEDLFNPCQAIKC-KHGKCRLSGLGQP-Y-CECSSGYTGDS 1442

Db   1263 CENSYSMCSAN 1273
          |:  | | ::
```

```
Qy   1443 CDREIS-CRGE 1452


RESULT   13
ID   O13149      PRELIMINARY;      PRT;  2447 AA.
AC   O13149;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH 2 (FRAGMENT).
OS   FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   TETRAODONTIFORMES; TETRAODONTOIDEI; TETRAODONTIDAE; FUGU.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NAKAMURA T., TROWSDALE J.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB004829; D1021371; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 20.
     PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   2447 AA;  262542 MW;  3CDA4F7A CRC32;

  Query Match           5.9%;  Score 672;  DB 13;  Length 2447;
  Best Local Similarity  29.1%;  Pred. No. 9.53e-103;
  Matches   163;  Conservative  115;  Mismatches 234;  Indels  48;  Gaps  43;

Db    332 DACISKPCKGGSKCDTNPISGMFNCNCPSGYTGSTCSIDRDECSIGTNPCEHGGQC-VN- 389
          ::|:|:|||  : |:::|:   : | || |: |  | :    | | :||| ||| | ::
Qy    916 NPCLSNPCKNDGTCNSDPVD-FYRCTCPYGFKGQDCDVPIHAC-I-SNPCKHGGTCHLKE 972

Db    390 TE-GSFTCNCAKGYAGPRCEQDVNECASNPCQNDGTCLDRIGDYSCICMPGFGGTHCENE 448
          |  :| | || |: |  || :|::|  | |:|::||:| | :|:|:| | : |  ||:
Qy    973 GEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEK 1032

Db    449 LNECL-S-SPCLNRGKCLDQVSRFVCECPAGFSGEMCQIDIDECSSTPCLNGAKCIDLPN 506
          |: |    :|| : :||:     | |:| :|: || |:||:|:|    | ||| | |  |
Qy   1033 LDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVN 1092

Db    507 GYDCECAEGFKGLLCEENINDCVPEPCHHGQCKDGIATFSCECYAGYTGAICNIQVQECH 566
          || | |:||: ||:||  ::  :  | : : | :     : :|      :||   :   :
Qy   1093 GYTCICPEGYSGLFCE--FSPPMVLP-RTSPCDNFDCQNGAQCIVRINEPICQC-LPGYQ 1148

     567 SNPCQNRGRCIDLVN--AYQCNCPPGISGVNCEINEDDCASNLCVYGECQDGINEYKCVC 624
          :: |: :   ::::|  :|    |::      :| : : :          | ::   |
Qy   1149 GEKCE-KLVSVNFINKESYL-QIPSAKVRPQTNITLQIATDEDSGILLYK-GDKDHIAV- 1204

Db    625 SPGYTGDKCDVDINECSSNPCMSGGTCVDNVN-G-FHCLCPPSTYGLLCLSGTDHCVAQP 682
            | | :      : :|:|  |:   |: :| | || :   :    | ||  |   : |
Qy   1205 EL-YRG-RVRAS-YDTGSHP-ASAIYSVETINDGNFHIVELLALDQSLSLS-VD--GGNP 1257

Db    683 CVHGKCIEQQNGYFCQCEAGWVGQHCEQEKDECL-PNPCQNGGSCLD-RHNGFTCVCQAG 740
          :  :   |    | :     ||      :   |   | ||| |     :| :
Qy   1258 KIITNLSKQSTLNF-DSPL-YVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQD 1315

Db    741 YRGVNCEKNI--D-E-CTSGPCLNQGICI-DGLNSYTCQCVPPFAGEHCEVEL-DPCSSR 794
          :: |  :  |    | |    | :| |   : ::||:|     |  |:    ||| :
Qy   1316 FQKVPMQTGILPGCEPCHKKVC-AHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGN 1374

Db    795 PCQRGGVCLPSADYTYFTCRCPAGWQGLHCSE--DV-NECKKNPCRNGGHC-INSPG-SY 849
          | :|  |||   ::| :|:|  |  |: | |  |: | |    |::|  | ::: | :|
Qy   1375 KCVHG-TCLPINAFSY-SCKCLEGHGGVLCDEEEDLFNPCQAIKCKHG-KCRLSGLGQPY 1431

Db    850 ICKCPSGYSGHNCQTDIDDC 869
          | |:|||:| :|: :|  |
Qy   1432 -CECSSGYTGDSCDREIS-C 1449


RESULT   14
ID   O57409      PRELIMINARY;      PRT;  615 AA.
AC   O57409;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAB.
GN   DELTAB.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98165391.
RA   HADDON C., SMITHERS L., SCHNEIDER-MAUNOURY S., COCHE T., HENRIQUE D.,
RA   LEWIS J.;
RT   "Multiple delta genes and lateral inhibition in zebrafish primary
RT   neurogenesis.";
RL   DEVELOPMENT 125:359-370(1998).
DR   EMBL; AF006488; G2772825; -.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   615 AA;  67592 MW;  FC6348F8 CRC32;

  Query Match           5.8%;  Score 657;  DB 13;  Length 615;
  Best Local Similarity  38.1%;  Pred. No. 1.00e-99;
  Matches   91;  Conservative  56;  Mismatches  76;  Indels  16;  Gaps  11;

Db    274 CTNHKPCANGATCTNTGQGSYTCTCRPGFGGTNCELEINECDCNPCKNGGSCN--DLEND 331
          | :: || | :|| :    | |||  || | :|:: |: |  ||||:||:|:  : |:|
Qy    918 CLSN-PCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEED 976

Db    332 -YSCTCPQGFYGKNCEIIAMTCADDPCFNGGTCEEKFTGGYVCRCPPTFTGSNCEKRLDR 390
           : | |::|| | |||:    | |: | | :|| : :   | | ||| :||  || :||
Qy    977 GFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINN-YTCLCPPEYTGELCEEKLDF 1035

Db    391 CSH--KPCANGGECVDLGASAL-CRCRPGFSGSRCETNIDDCARYPCQNAGTCQDGINDY 447
          |::  :|| : : |: | : :: | | ||: | :|: ::|||    | |:: | |::| |
Qy   1036 CAQDLNPCQHDSKCI-LTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGY 1094

Db    448 TCTCTLGFTGKNCSLRADACL--TNPC--L--HGGT-CFTHFSGPVCQCVPGFMGSTCE 499
          || |  |::|  | : :   |  |:||  :  : |: |: ::: |:|||:||: |  ||
Qy   1095 TCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCE 1153


RESULT   15
ID   O42347      PRELIMINARY;      PRT;  1212 AA.
AC   O42347;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C-SERRATE-2 (FRAGMENT).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97184054.
RA   HAYASHI H., MOCHII M., KODAMA R., HAMADA Y., MIZUNO N., EGUCHI G.,
RA   TACHI C.;
RT   "Isolation of a novel chick homolog of Serrate and its coexpression
RT   with C-Notch-1 in chick development.";
RL   INT. J. DEV. BIOL. 40:1089-1096(1996).
DR   EMBL; D87558; D1022568; -.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   1212 AA;  134188 MW;  0ECF076C CRC32;
```

Query Match 5.8%; Score 659; DB 13; Length 1212;
Best Local Similarity 39.6%; Pred. No. 3.96e-100;
Matches 93; Conservative 48; Mismatches 79; Indels 15; Gaps 10;

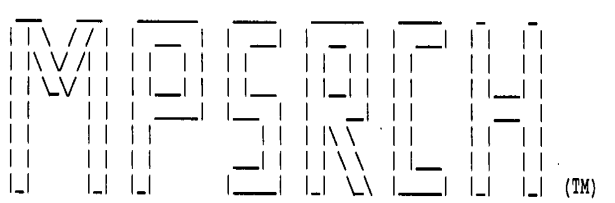
```
Db    285 HPCLNGGTCMNTEPDE-YRCACPDGYSGKNCEIAEHACVSNPCANGGTCH--E-ISSSFK 340
          :|| | ||| |::| : |||:|| |: | :|::: |||:|||| :|||||  |    :|
Qy    921 NPCKNDGTC-NSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEEDGFW 979

Db    341 CHCPSGWSGPTCAIDIDECASNPCAQGGTCIDHINSFECICPQQWIGATCQLDANEC-EG 399
          | |: |  |  | :::|:|  | |   :||:| ||:: |:|| :  |  |:   : | :
Qy    980 CICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFCAQD 1039

Db    400 -KPCVNAYSCKNLIGGYYCDCIPGWKGVNCHININDCHGQ-CQHGGTCKDEVNDYHCICP 457
           :|| :   |     |: ||| ||  | :| |:::||:   | :|: | | || | ||||
Qy   1040 LNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICP 1099

Db    458 RGFTGKNCE-----I-E-TNECESNPCQNGGRCKDLVNGFTCLCAQGFSGVFCEM 505
          |::|  ||     :   |: |::  ||||::|   :|   | |  |: |  ||:
Qy   1100 EGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGEKCEK 1154
```

Search completed: Fri May 28 08:35:44 1999
Job time : 261 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:41:25 1999; MasPar time 6.83 Seconds
326.882 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-3
Description: (1-105) from US09191647.pep
Perfect Score: 730
Sequence: 1 SPCTCSNNIVDCRGKGLMEI.....ITEIAKGLFDGLVSLQLLLL 105

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 28.979; Variance 136.917; scale 0.212

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	581	79.6	1534	30	W46966	Amino acid sequence o	2.48e-40
2	459	62.9	1480	5	R25079	Drosophila SLIT prote	1.22e-29
3	234	32.1	345	23	W09405	Pineal gland specific	1.95e-10
4	224	30.7	692	2	R08038	Rat testicular lutein	1.30e-09
5	223	30.5	332	15	R87953	Bovine neurotrophic b	1.57e-09
6	223	30.5	368	1	R05159	Sequence of human bon	1.57e-09
7	221	30.3	369	15	R87951	Rat neurotrophic bigl	2.30e-09
8	221	30.3	369	15	R87952	Human neurotrophic bi	2.30e-09
9	221	30.3	634	6	R30520	N-terminal of LH rece	2.30e-09
10	221	30.3	689	6	R30509	N-terminal of LH rece	2.30e-09
11	221	30.3	692	6	R30503	N-terminal of LH rece	2.30e-09
12	221	30.3	695	6	R30506	N-terminal of LH rece	2.30e-09
13	221	30.3	695	6	R30525	N-terminal of LH rece	2.30e-09
14	221	30.3	695	6	R30524	N-terminal of LH rece	2.30e-09
15	221	30.3	696	6	R30523	N-terminal of LH rece	2.30e-09
16	221	30.3	698	6	R30505	N-terminal of LH rece	2.30e-09
17	219	30.0	690	6	R30514	N-terminal of LH rece	3.35e-09
18	219	30.0	693	6	R30510	N-terminal of LH rece	3.35e-09
19	219	30.0	696	6	R30521	N-terminal of LH rece	3.35e-09
20	219	30.0	696	6	R30526	N-terminal of LH rece	3.35e-09
21	219	30.0	696	6	R30513	N-terminal of LH rece	3.35e-09
22	219	30.0	699	6	R30515	N-terminal of LH rece	3.35e-09
23	217	29.7	390	20	W06532	Gonadotropin receptor	4.89e-09
24	217	29.7	695	21	W14782	FSH receptor.	4.89e-09
25	217	29.7	695	8	R42082	FSH receptor.	4.89e-09
26	216	29.6	196	5	R29102	Drosophila SLIT prote	5.91e-09
27	216	29.6	695	5	R27558	FSHR.	5.91e-09
28	214	29.3	1091	27	W41641	Sequence used in dete	8.61e-09
29	211	28.9	620	6	R30522	N-terminal of LH rece	1.51e-08
30	211	28.9	696	6	R30519	N-terminal of LH rece	1.51e-08
31	209	28.6	904	39	W86351	Human DNAX toll-like	2.20e-08
32	198	27.1	605	17	R85888	WD-40 domain-contg. i	1.72e-07
33	193	26.4	837	39	W86361	Human DNAX toll-like	4.37e-07
34	190	26.0	560	12	R71294	Human glycoprotein V.	7.63e-07
35	188	25.8	139	8	R42263	Decorin sequence PT-7	1.10e-06
36	188	25.8	186	8	R42264	Decorin sequence PT-7	1.10e-06
37	188	25.8	234	8	R42265	Decorin sequence PT-7	1.10e-06
38	188	25.8	280	8	R42266	Decorin sequence PT-7	1.10e-06
39	188	25.8	305	8	R42267	Decorin sequence PT-7	1.10e-06
40	188	25.8	331	8	R42260	Mature decorin PT-65.	1.10e-06
41	188	25.8	342	17	R89439	Human recombinant dec	1.10e-06
42	188	25.8	353	1	R05160	Sequence of human bon	1.10e-06
43	188	25.8	1388	18	R89471	Collagen/decorin fusi	1.10e-06
44	187	25.6	603	17	R85889	WD-40 domain-contg. r	1.33e-06
45	182	24.9	799	39	W86352	Human DNAX toll-like	3.35e-06

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 79.6%; Score 581; DB 30; Length 1534;
Best Local Similarity 72.8%; Pred. No. 2.48e-40;
Matches 75; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

Db 286 ctcsngivdcrgkgltaipanlpetmteirlelngiksippgafspyrklrridlsnnqi 345
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 346 aeiapdafqglrslnslvlygnkitdlprgvfgglytlqllll 388

```
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match          62.9%;  Score 459;  DB 5;  Length 1480;
  Best Local Similarity  53.8%;  Pred. No. 1.22e-29;
  Matches   56;  Conservative   24;  Mismatches  24;  Indels   0;  Gaps   0;

Db    298 pcrcadgivdcreksltsvpvtlpddttdvrleqnfitelppksfssfrrlrridlsnnn 357
          || |:: ||||| |:|  :|  ||:   ::||||| |  :|: :|: :::|:|||:|:|
Qy      2 PCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQ 61

Db    358 isriahdalsglkqlttlvlygnkikdlpsgvfkglgslrllll 401
          || || ||: ||| ||:|||||||| ::: |:| || ||:||||
Qy     62 ISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    3
ID   W09405 standard; Protein; 345 AA.
AC   W09405;
DT   17-SEP-1997  (first entry)
DE   Pineal gland specific gene-1 protein.
KW   PGSG-1; pineal gland; epiphysis cerebri; tumour; precocious puberty;
KW   hydrocephalus; papilledema; intracranial pressure; circadian rhythm;
KW   pituitary secretion; luteinising hormone; growth hormone;
KW   follicular stimulating hormone.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..21
FT                   /note= "Putative leader sequence"
FT   protein         22..344
FT                   /label= Pineal_gland_specific_gene-1
FT   region          22..283
FT                   /note= "Putative soluble portion of the protein"
FT   region          284..344
FT                   /note= "Insoluble portion of the protein"
FT   region          283..344
FT                   /note= "Putative transmembrane portion of the protein"
PN   WO9639158-A1.
PD   12-DEC-1996.
PF   05-JUN-1995; U07067.
PR   05-JUN-1995; WO-U07067.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   He WW,  Rosen CA;
DR   WPI; 97-042840/04.
DR   N-PSDB; T47647.
PT   Pineal gland specific gene-1 and corresponding protein - used in the
PT   treatment of pineal tumours and alleviation of side effect, e.g.
PT   precocious puberty, hydrocephalus etc.
PS   Claim 1; Page 42-43; 56pp; English.
CC   The present sequence represents a novel isolated polypeptide,
CC   pineal gland specific gene-1 (PGSG-1), which was derived from a human
CC   pineal gland tissue cDNA library. The PGSG-1 polypeptide may be used
CC   to treat pineal tumours and thereby treat the side effects, including
CC   precocious puberty, hydrocephalus, papilledema and other signs of
CC   increased intracranial pressure. PGSG-1 and its protein may be used to
CC   regulate biological rhythms, in particular circadian rhythms, and to
CC   regulate pituitary secretion of hormones which regulate the onset of
CC   puberty, e.g. luteinising hormone, follicular stimulating hormone (FSH)
CC   and growth hormone (GH) released by the pituitary. The (ant)agonists
CC   which act against the protein may also be used to regulate the secretion
CC   of these hormones. The PGSG-1 gene and proteins may also be used to
CC   diagnose a mutation in the PGSG-1 gene and hence susceptibility to a
CC   disease mentioned above.
SQ   Sequence   345 AA;

  Query Match          32.1%;  Score 234;  DB 23;  Length 345;
  Best Local Similarity  38.8%;  Pred. No. 1.95e-10;
  Matches   40;  Conservative   21;  Mismatches  42;  Indels   0;  Gaps   0;

Db     27 cqsstnfvdcsqqglaeipshlppqtrtlhlqdnqihhlpafafrsvpwlmtlnlsnnsl 86
          | | |:||   || |||::||    ::|::| |  :|| ||     |: :::|:| :
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     87 snlapgafhglqhlqvlnltqnsllslesrlfhslpqlreldl 129
          . |::|| ||:||  |  | |  | :  :   || :|  |: | |
```

```
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT   4
ID   R08038 standard; protein; 692 AA.
AC   R08038;
DT   26-FEB-1991  (first entry)
DE   Rat testicular luteinising hormone/choriogonadotropin receptor.
KW   LH/CG receptor; FSH receptor; TSH receptor; fertility; breast cancer;
KW   prostate cancer;thyroid cancer; osteoporosis; Graves disease;
KW   polycistronic ovarian disease; vasomoter instability.
OS   Rattus rattus.
PN   WO9013643-A.
PD   15-NOV-1990.
PF   04-MAY-1990; U02488.
PR   05-MAY-1989; US-347683.
PA   (GETH ) GENENTECH INC.
     Nikolics K, McFarland KC, Segaloff DL, Seeburg PH;
     WPI; 90-361478/48.
DR   N-PSDB; Q06634
PT   Pharmaceutical compsn. contg. hormone receptor mol - used for
PT   treating fertility, breast-and prostate-cancer and osteoporosis,
PT   etc.
PS   Disclosure; fig 6; 78pp; English.
CC   This rat testicular follicle-stimulating hormone (FSH) receptor.
CC   This receptor is useful in a pharmaceutical compsn. for treating
CC   e.g. breast-, prostate- and thyroid cancer, fertility, osteopor-
CC   osis, vasomoter instability and polycistronic ovarian disease.
CC   LH/CG- or TSH-receptors can also be used, to treat e.g. Graves
CC   disease.  Abs can be used to inhibit receptor binding and for imag-
CC   ing and therapy.  See also R08015-23, R08035-36 and Q06633.
SQ   Sequence   692 AA;

  Query Match          30.7%;  Score 224;  DB 2;  Length 692;
  Best Local Similarity  32.4%;  Pred. No. 1.30e-09;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
         | ||| :  |:   : |||::||   :|:|:    :: || |:|: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 levieadvfsnlpklheiriekannvlyinpeafqnlpslrylli 127
          : | :| |  | |  : :  :|::  |    |: | ||: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


ULT   5
R87953 standard; Protein; 332 AA.
AC   R87953;
DT   20-MAR-1996  (first entry)
DE   Bovine neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   region          7..23
FT                   /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s) -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 3; Fig 8; 60pp; English.
CC   Bovine biglycan (R87953) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  It can be used to enhance
CC   the survival and maintain the structure and function of CNS neurons
CC   during normal ageing as well as after pathological and/or traumatic
CC   nervous system damage.  It can also be used to restore function
CC   following nervous system lesions and degenerative diseases, and to
CC   improve learning efficiency and memory in the elderly and in patients
CC   with dementia.
SQ   Sequence   332 AA;

  Query Match          30.5%;  Score 223;  DB 15;  Length 332;
  Best Local Similarity  38.6%;  Pred. No. 1.57e-09;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    114 lveippnlpsslvelrihdnrirkvpkgvfsglrnmneiemggnplensgfepgafdglk 173
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    174 -lnylriseakltgipkdlpetlnelhl 200
           | | :  |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   6
ID   R05159 standard; protein; 368 AA.
AC   R05159;
DT   09-OCT-1990  (first entry)
DE   Sequence of human bone proteoglycan I (biglycan).
KW   Osteoporosis; rheumatoid arthritis; Paget's disease;
KW   atherosclerosis; periodontal; human bone matrix; proteoglycan.
OS   Homo sapiens.
PN   US7432044-A.
PD   17-APR-1990.
PF   3-NOV-1989; 432044.
PR   3-NOV-1989; US-432044.
PA   (USSH) Nat Inst of Health.
PI   Termine J;
DR   WPI; 90-178641/23.
DR   N-PSDB; Q04490.
PT   Human bone matrix DNA and proteins -
PT   used in detection, diagnosis and treatment involving skeletal
PT   and/or connective tissue disease states.
PS   Disclosure; p; English.
CC   Probes and Abs raised to the proteins can be used to determine
CC   their levels useful in diagnosis of associated conective tissue
CC   diseases states such as osteoporosis, osteo/rheumatoid arthritis,
CC   Paget's disease, artherosclerosis and periodontal disease.
CC   Proteins may also be used to induce or block biological function.
SQ   Sequence   368 AA;

  Query Match          30.5%;  Score 223;  DB 1;  Length 368;
  Best Local Similarity  37.5%;  Pred. No. 1.57e-09;
  Matches   33;  Conservative   26;  Mismatches  26;  Indels   3;  Gaps   2;

Db    150 lveippnlpsslvdvrihdnrirkvpkgvfsglrnmnciemggnplensgfepgafdglk 209
          |:|||:||| ::|::|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    210 -lnylriseakltgipkdlpetlnelhl 236
           | | :  |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   7
ID   R87951 standard; Protein; 369 AA.
AC   R87951;
DT   20-MAR-1996  (first entry)
DE   Rat neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   peptide         1..37
FT                   /label= Sig_peptide
FT   region          44..60
```

```
FT                  /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
DR   N-PSDB; T08768.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)  -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 1; Page 44-45; 60pp; English.
CC   Rat biglycan (R87951) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  Recombinant biglycan,
CC   obtd. by expression of encoding cDNA (T08768) in eukaryotic host
CC   cells, can be used to enhance the survival and maintain the structure
CC   and function of CNS neurons during normal ageing as well as after
CC   pathological and/or traumatic nervous system damage.  It can also
CC   be used to restore function following nervous system lesions and
CC   degenerative diseases, and to improve learning efficiency and memory
CC   in the elderly and in patients with dementia.
SQ   Sequence   369 AA;

  Query Match           30.3%;  Score 221;  DB 15;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 2.30e-09;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 lveippnlpsslvelrihdnrirkvpkgvfsglrnmnciemggnplensgfepgafdglk 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -lnylriseakltgipkdlpetlnelhl 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    8
ID   R87952 standard; Protein; 369 AA.
AC   R87952;
DT   20-MAR-1996  (first entry)
DE   Human neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..37
                     /label= Sig_peptide
     region          44..60
FT                   /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)  -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 3; Fig 8; 60pp; English.
CC   Human biglycan (R87952) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  It can be used to enhance
CC   the survival and maintain the structure and function of CNS neurons
CC   during normal ageing as well as after pathological and/or traumatic
CC   nervous system damage.  It can also be used to restore function
CC   following nervous system lesions and degenerative diseases, and to
CC   improve learning efficiency and memory in the elderly and in patients
CC   with dementia.
SQ   Sequence   369 AA;

  Query Match           30.3%;  Score 221;  DB 15;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 2.30e-09;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 lveippnlpsslvelrihdnrirkvpkgvfsglrnmnciemggnplensgfepgafdglk 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -lnylriseakltgipkdlpetlnelhl 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    9
ID   R30520 standard; protein; 634 AA.
AC   R30520;
DT   10-MAY-1993  (first entry)
DE   N-terminal of LH receptor/FSH receptor chimaera #29.
KW   Follicle stimulating hormone receptor; luteinising hormone receptor;
KW   human chorionic gonadotrophin; glycoprotein hormone receptor;
KW   chimaera; chimera.
OS   Chimaeric; homo sapiens.
PN   WO9222667-A.
PD   23-DEC-1992.
PF   12-JUN-1992; U04987.
PR   14-JUN-1991; US-715911.
PA   (UYNE-) UNIV NEW JERSEY.
PI   Bernard M,  Moyle WR,  Myers R;
DR   WPI; 93-018150/02.
PT   Glyco:protein hormone receptor analogues - having binding
PT   affinity to human chorionic gonadotrophin, luteinising and
PT   follicle stimulating hormones, useful in bio:immunoassays
PS   Examples; Fig 12; 103pp; English.
CC   This sequence represents the N-terminal of a novel protein having a
CC   binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC   hormone (LH), and follicle stimulating hormone (FSH). The protein
CC   itself is a chimaera having residues from both  thew FSH receptor,
CC   and LH receptor. The receptor analogues can be used in bioimmunoassays
CC   for the simultaneous detection of both LH  (or hCG) and FSH as
CC   well as their ratio of biological activities. The analogues can also
CC   be used for raising, purifying and assaying antibodies to the
CC   analogues. Coding sequence for the chimaera was produced by two step
CC   PCR.
SQ   Sequence   634 AA;

  Query Match           30.3%;  Score 221;  DB 6;  Length 634;
  Best Local Similarity  32.4%;  Pred. No. 2.30e-09;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
          | ||| :  |:   : |||::||   :|:|:    :: || |:|: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
           : | :| |  |  |  : :  :|::  |    |: | ||: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT   10
ID   R30509 standard; protein; 689 AA.
AC   R30509;
DT   10-MAY-1993  (first entry)
DE   N-terminal of LH receptor/FSH receptor chimaera #18.
KW   Follicle stimulating hormone receptor; luteinising hormone receptor;
KW   human chorionic gonadotrophin; glycoprotein hormone receptor;
KW   chimaera; chimera.
OS   Chimaeric; homo sapiens.
PN   WO9222667-A.
PD   23-DEC-1992.
PF   12-JUN-1992; U04987.
PR   14-JUN-1991; US-715911.
PA   (UYNE-) UNIV NEW JERSEY.
```

PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 689 AA;

Query Match 30.3%; Score 221; DB 6; Length 689;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105

RESULT 11
ID R30503 standard; protein; 692 AA.
AC R30503;
DT 10-MAY-1993 (first entry)
DE N-terminal of LH receptor/FSH receptor chimaera #10.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimaera; chimera.
OS Chimaeric; homo sapiens.
PN WO9222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (UYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 692 AA;

Query Match 30.3%; Score 221; DB 6; Length 692;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105

RESULT 12
ID R30506 standard; protein; 695 AA.
AC R30506;
DT 10-MAY-1993 (first entry)
DE N-terminal of LH receptor/FSH receptor chimaera #15.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimaera; chimera.
OS Chimaeric; homo sapiens.
PN WO9222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (UYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 695 AA;

Query Match 30.3%; Score 221; DB 6; Length 695;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105

RESULT 13
ID R30525 standard; protein; 695 AA.
AC R30525;
DT 10-MAY-1993 (first entry)
DE N-terminal of LH receptor/FSH receptor chimaera #34.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimaera; chimera.
OS Chimaeric; homo sapiens.
PN WO9222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (UYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays

CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 695 AA;

Query Match 30.3%; Score 221; DB 6; Length 695;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105

RESULT 14
R30524 standard; protein; 695 AA.
AC R30524;
DT 10-MAY-1993 (first entry)
DE N-terminal of LH receptor/FSH receptor chimaera #33.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimaera; chimera.
OS Chimaeric; homo sapiens.
PN WO9222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (UYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 695 AA;

Query Match 30.3%; Score 221; DB 6; Length 695;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105

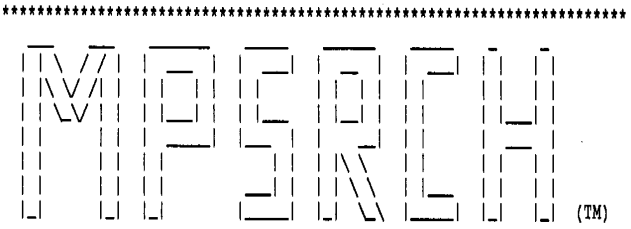
RESULT 15
ID R30523 standard; protein; 696 AA.
AC R30523;
DT 10-MAY-1993 (first entry)
DE N-terminal of LH receptor/FSH receptor chimaera #32.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimaera; chimera.
OS Chimaeric; homo sapiens.
PN WO9222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (UYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimaera having residues from both thew FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimaera was produced by two step
CC PCR.
SQ Sequence 696 AA;

Query Match 30.3%; Score 221; DB 6; Length 696;
Best Local Similarity 32.4%; Pred. No. 2.30e-09;
Matches 34; Conservative 27; Mismatches 42; Indels 2; Gaps 2;

Db 23 chcsnrvflcqdskvteiptdlprnaielrfvltklrvipkgsfagfgdlekieisqndv 82
Qy 3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db 83 levieadvfsnlpklheiriekannllyinpeafqnlpslrylli 127
Qy 63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105

Search completed: Fri May 28 08:42:12 1999
Job time : 47 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

ssrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:42:30 1999; MasPar time 7.20 Seconds
584.123 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-3
Description: (1-105) from US09191647.pep
Perfect Score: 730
Sequence: 1 SPCTCSNNIVDCRGKGLMEI.....ITEIAKGLFDGLVSLQLLLL 105

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 42.417; Variance 97.098; scale 0.437

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	459	62.9	1469	2	B36665	slit protein 2 precur	1.55e-57
2	459	62.9	1480	2	A36665	slit protein 1 precur	1.55e-57
3	223	30.5	694	2	JC4301	follitropin receptor	3.26e-18
4	221	30.3	368	1	BGHUN	biglycan precursor -	6.69e-18
5	221	30.3	369	2	S32793	biglycan precursor -	6.69e-18
6	221	30.3	369	2	S20811	proteoglycan I - mous	6.69e-18
7	221	30.3	369	2	S32559	biglycan precursor -	6.69e-18
8	221	30.3	692	2	A34548	follitropin receptor	6.69e-18
9	221	30.3	694	2	JC2237	follitropin receptor,	6.69e-18
10	220	30.1	695	2	I45896	follicle stimulating	9.58e-18
11	219	30.0	695	2	JC1493	follitropin receptor	1.37e-17
12	218	29.9	695	2	JN0898	follitropin receptor	1.96e-17
13	217	29.7	695	1	QRHUFT	follitropin receptor	2.80e-17
14	214	29.3	1091	2	A58532	glial cell membrane g	8.16e-17
15	209	28.6	313	2	G02020	p37NB - human	4.81e-16
16	204	27.9	357	2	S24317	decorin precursor - c	2.81e-15
17	200	27.4	605	2	JC5239	insulin-like growth f	1.15e-14
18	200	27.4	662	2	S42799	garp precursor - huma	1.15e-14
19	198	27.1	605	2	A41915	insulin-like growth f	2.31e-14
20	197	27.0	354	2	S29145	decorin precursor - r	3.27e-14
21	193	26.4	360	2	S06280	decorin precursor - b	1.32e-13
22	191	26.2	360	2	I47020	decorin - rabbit	2.64e-13
23	190	26.0	560	2	A60164	platelet membrane gly	3.73e-13
24	188	25.8	359	1	NBHUC8	decorin precursor - h	7.44e-13
25	187	25.6	603	2	JC1282	insulin-like growth f	1.05e-12
26	186	25.5	354	2	A55454	decorin precursor - m	1.48e-12
27	184	25.2	322	2	S72271	proteoglycan Lb precu	2.95e-12
28	184	25.2	361	2	A53860	chondroadherin precur	2.95e-12
29	182	24.9	603	2	JC6128	insulin-like growth f	5.85e-12
30	180	24.7	316	2	A41781	proteoglycan-Lb - chi	1.16e-11
31	179	24.5	382	2	I39068	proline- arginine-ric	1.63e-11
32	175	24.0	1115	2	S40241	G protein-coupled rec	6.33e-11
33	173	23.7	907	2	JE0176	orphan G protein-coup	1.24e-10
34	170	23.3	536	2	A34901	lysine carboxypeptida	3.41e-10
35	170	23.3	1134	1	A29944	chaoptin precursor -	3.41e-10
36	155	21.2	682	2	A49121	cell-surface molecule	4.87e-08
37	155	21.2	682	2	A43318	connectin precursor -	4.87e-08
38	153	21.0	1535	2	S46224	peroxidasin - fruit f	9.34e-08
39	152	20.8	626	1	NBHUIA	platelet glycoprotein	1.29e-07
40	148	20.3	343	2	A41748	lumican precursor - c	4.69e-07
41	148	20.3	375	2	S05390	fibromodulin precurso	4.69e-07
42	148	20.3	440	2	A39613	oligodendrocyte-myeli	4.69e-07
43	148	20.3	440	2	A47530	oligodendrocyte-myeli	4.69e-07
44	145	19.9	1025	1	A57676	protein kinase Xa21 (	1.22e-06
45	142	19.5	176	1	A46606	platelet glycoprotein	3.17e-06

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           62.9%;  Score 459;  DB 2;  Length 1469;
  Best Local Similarity  53.8%;  Pred. No. 1.55e-57;
  Matches    56;  Conservative   24;  Mismatches  24;  Indels   0;  Gaps   0;

Db    298 PCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRIDLSNNN 357
          || |:: ||||| |:|  :|  ||:   ::||||| |  :|: :|: :::|:|||:|:|
Qy      2 PCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQ 61

Db    358 ISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
          || || ||: ||| ||:|||||||| ::: |:| || ||:||||
Qy     62 ISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
   #authors     Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
   #journal     Genes Dev. (1990) 4:2169-2187
   #title       slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
   #cross-references MUID:91099665
   #accession   A36665
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-1480 ##label ROT
      ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
   #gene        FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing

FEATURE
   66-91        #domain proteoglycan amino-terminal homology #label
                  PAH1\
   101-124      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
   125-148      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
   149-172      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
   173-196      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
   197-220      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
   228-272      #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
   288-313      #domain proteoglycan amino-terminal homology #label
                  PAH2\
   323-346      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
   347-370      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
   371-394      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494      #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537      #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695      #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733      #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   791-814      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
   815-838      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
   846-890      #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061    #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           62.9%;  Score 459;  DB 2;  Length 1480;
  Best Local Similarity  53.8%;  Pred. No. 1.55e-57;
  Matches    56;  Conservative   24;  Mismatches  24;  Indels   0;  Gaps   0;

Db    298 PCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRIDLSNNN 357
          || |:: ||||| |:|  :|  ||:   ::||||| |  :|: :|: :::|:|||:|:|
Qy      2 PCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQ 61

Db    358 ISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
          || || ||: ||| ||:|||||||| ::: |:| || ||:||||
Qy     62 ISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    3
ENTRY           JC4301     #type complete
TITLE           follitropin receptor - pig
ALTERNATE_NAMES  follicle-stimulating hormone receptor
```

```
ORGANISM        #formal_name Sus scrofa domestica #common_name domestic pig
DATE            16-Nov-1995 #sequence_revision 08-Feb-1996 #text_change
                  17-Mar-1999
ACCESSIONS      JC4301
REFERENCE       JC4301
  #authors      Remy, J.J.; Lahbib-Mansais, Y.; Yerle, M.; Bozon, V.;
                  Couture, L.; Pajot, E.; Grebert, D.; Salesse, R.
  #journal      Gene (1995) 163:257-261
  #title        The porcine follitropin receptor: cDNA cloning, functional
                  expressionand chromosomal localization of the gene.
  #cross-references MUID:96011644
  #accession    JC4301
    ##molecule_type mRNA
    ##residues      1-694 ##label REM
    ##cross-references GB:L31966
    ##experimental_source ovarian granulosa cells
COMMENT    This receptor belongs to the family of the G-protein coupled
             receptors. It has the functional roles of high affinity for
             follicle-stimulating hormone. It also plays an essential role in
             spermatogenesis in male and oogenesis in female.
GENETICS
  #gene         fshr
  #map_position 3 q2.2-q2.3
CLASSIFICATION  #superfamily glycoprotein hormone receptor; leucine-rich
                  alpha-2-glycoprotein repeat homology
KEYWORDS        G protein-coupled receptor; hormone receptor; transmembrane
                  protein
FEATURE
  1-365             #domain follicle-stimulating hormone binding #status
                      predicted #label HOB\
  366-388           #domain transmembrane #status predicted #label TM1\
  398-420           #domain transmembrane #status predicted #label TM2\
  443-464           #domain transmembrane #status predicted #label TM3\
  485-507           #domain transmembrane #status predicted #label TM4\
  528-549           #domain transmembrane #status predicted #label TM5\
  573-596           #domain transmembrane #status predicted #label TM6\
  608-629           #domain transmembrane #status predicted #label TM7
SUMMARY         #length 694  #molecular-weight 78278  #checksum 409

 Query Match          30.5%;  Score 223;  DB 2;  Length 694;
 Best Local Similarity  33.3%;  Pred. No. 3.26e-18;
 Matches   35;  Conservative   27;  Mismatches  41;  Indels   2;  Gaps   2;

Db     22 CHCSNGVFLCQESKVTEIPPDLPRNAVELRFVLTKLRAIPKGAFSGFGDLEKIEISQNDV 81
          | ||| :  |:   : |||::||   ||:|:    ::||| |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     82 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 126
          : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT    4
ENTRY           BGHUN      #type complete
TITLE           biglycan precursor - human
ALTERNATE_NAMES  cartilage proteoglycan protein I; dermatan sulfate
                  proteoglycan I (DS-PGI); proteoglycan I core protein (PG-I)
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            21-Apr-1992 #sequence_revision 26-May-1995 #text_change
                  14-Nov-1997
ACCESSIONS      A40757; A32458; S05639; A28457; S14349
REFERENCE       A40757
  #authors      Fisher, L.W.; Heegaard, A.M.; Vetter, U.; Vogel, W.; Just,
                  W.; Termine, J.D.; Young, M.F.
  #journal      J. Biol. Chem. (1991) 266:14371-14377
  #title        Human biglycan gene. Putative promoter, intron-exon
                  junctions, and chromosomal localization.
  #cross-references MUID:91317791
  #accession    A40757
    ##molecule_type DNA
    ##residues      1-368 ##label FIS
    ##cross-references GB:M65151; GB:M65152; GB:M65153
REFERENCE       A32458
  #authors      Fisher, L.W.; Termine, J.D.; Young, M.F.
  #journal      J. Biol. Chem. (1989) 264:4571-4576
  #title        Deduced protein sequence of bone small proteoglycan I
                  (Biglycan) shows homology with proteoglycan II (Decorin)
                  and several nonconnective tissue proteins in a variety of
                  species.
  #cross-references MUID:89174714
  #accession    A32458
    ##molecule_type mRNA
    ##residues      1-138,'NV',141-162,'DV',165-368 ##label FI2
    ##cross-references GB:J04599
    ##note          parts of this sequence, including the amino end of the
                      mature protein, were determined by protein sequencing
REFERENCE       S05639
  #authors      Roughley, P.J.; White, R.J.
  #journal      Biochem. J. (1989) 262:823-827
  #title        Dermatan sulphate proteoglycans of human articular cartilage.
                  The properties of dermatan sulphate proteoglycans I and II.
  #cross-references MUID:90073579
  #accession    S05639
    ##molecule_type protein
    ##residues      38-41,'X',43-46,'X',48-57 ##label ROU
REFERENCE       A92656
  #authors      Fisher, L.W.; Hawkins, G.R.; Tuross, N.; Termine, J.D.
  #journal      J. Biol. Chem. (1987) 262:9702-9708
  #title        Purification and partial characterization of small
                  proteoglycans I and II, bone sialoproteins I and II, and
                  osteonectin from the mineral compartment of developing
                  human bone.
  #cross-references MUID:87250639
  #accession    A28457
    ##molecule_type protein
    ##residues      38-41,'X',43-62,'X',64-66 ##label FI3
    ##experimental_source bone
REFERENCE       S14349
  #authors      Stoecker, G.; Meyer, H.E.; Wagener, C.; Greiling, H.
  #journal      Biochem. J. (1991) 274:415-420
  #title        Purification and N-terminal amino acid sequence of a
                  chondroitin sulphate/dermatan sulphate proteoglycan
                  isolated from intima/media preparations of human aorta.
  #cross-references MUID:91174749
  #accession    S14349
    ##molecule_type protein
    ##residues      38-57 ##label STO
    ##experimental_source aorta
GENETICS
  #gene         GDB:BGN
    ##cross-references GDB:119727; OMIM:301870
  #map_position Xq28-Xq28
  #introns      80/1; 117/3; 189/1; 226/1; 257/2; 303/3
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        chondroitin sulfate proteoglycan; dermatan sulfate;
                  duplication; extracellular matrix; glycoprotein; tandem
                  repeat
FEATURE
  1-16              #domain signal sequence #status predicted #label SIG\
  17-37             #domain propeptide #status predicted #label PRO\
  38-368            #product biglycan #status predicted #label MAT\
  57-81             #domain proteoglycan amino-terminal homology #label PAH\
  91-114            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  115-138           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  139-159           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  160-183           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  184-207           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  209-229           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
```

```
   230-253          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   254-277          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   278-300          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   301-315          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   316-368          #domain proteoglycan carboxyl-terminal homology #label
                      PCH\
   42,47            #binding_site dermatan sulfate (Ser) (covalent) #status
                      experimental\
   180,198          #binding_site dermatan sulfate (Ser) (covalent) #status
                      predicted\
   270,311          #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY         #length 368  #molecular-weight 41654  #checksum 1684

  Query Match           30.3%;  Score 221;  DB 1;  Length 368;
  Best Local Similarity  38.6%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    150 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 209
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    210 -LNYLRISEAKLTGIPKDLPETLNELHL 236
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    5
ENTRY           S32793     #type complete
TITLE           biglycan precursor - rat
ALTERNATE_NAMES  dermatan sulfate proteoglycan I (DS-PGI); proteoglycan I core
                   protein (PG-I)
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change
                   29-Jan-1999
ACCESSIONS      S32793
REFERENCE       S32793
   #authors     Dreher, K.L.; Asundi, V.; Matzura, D.; Cowan, K.
   #journal     Eur. J. Cell Biol. (1990) 53:296-304
   #title       Vascular smooth muscle biglycan represents a highly conserved
                  proteoglycan within the arterial wall.
   #cross-references MUID:91184222
   #accession   S32793
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-369 ##label DRE
      ##cross-references GB:U17834; NID:g600497; PID:g600498
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        chondroitin sulfate proteoglycan; dermatan sulfate;
                  extracellular matrix; glycoprotein
FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-37            #domain propeptide #status predicted #label PRO\
   38-369           #product biglycan #status predicted #label MAT\
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH\
   42,48,181,199    #binding_site dermatan sulfate (Ser) (covalent) #status
                      predicted\
   271,312          #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY         #length 369  #molecular-weight 41706  #checksum 3056

  Query Match           30.3%;  Score 221;  DB 2;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    6
ENTRY           S20811     #type complete
TITLE           proteoglycan I - mouse
ALTERNATE_NAMES  biglycan
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   05-Dec-1997
ACCESSIONS      S20811; A57645; I49534
REFERENCE       S20811
   #authors     Naitoh, Y.; Suzuki, S.
   #submission  submitted to the EMBL Data Library, July 1990
   #description  Nucleotide sequences of cDNAs encoding mouse PGI and PGII.
   #accession   S20811
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-369 ##label NAI
      ##cross-references EMBL:X53928; NID:g53666; PID:g53667
REFERENCE       A57645
   #authors     Wegrowski, Y.; Pillarisetti, J.; Danielson, K.G.; Suzuki, S.;
                  Iozzo, R.V.
   #journal     Genomics (1995) 30:8-17
   #title       The murine biglycan: complete cDNA cloning, genomic
                  organization, promoter function, and expression.
   #cross-references MUID:96129295
   #accession   A57645
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-67,'W',69-369 ##label WEG
      ##cross-references GB:L20276; NID:g348961; PID:g348962
      ##note        authors translated the codon TGG for residue 58 as Cys
REFERENCE       I49534
   #authors     Rau, W.; Just, W.; Vetter, U.; Vogel, W.
   #journal     Mamm. Genome (1994) 5:395-396
   #title       A dinucleotide repeat in the mouse biglycan gene (EST) on the
                  X chromosome.
   #cross-references MUID:94319093
   #accession   I49534
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-67,'W',69-369 ##label RES
      ##cross-references GB:L20276; NID:g348961; PID:g348962
GENETICS
   #gene        Bgn
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
```

```
                repeat homology; proteoglycan amino-terminal homology;
                proteoglycan carboxyl-terminal homology
KEYWORDS        chondroitin sulfate proteoglycan; dermatan sulfate;
                extracellular matrix; glycoprotein

FEATURE
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH
SUMMARY         #length 369  #molecular-weight 41639  #checksum 3586

  Query Match           30.3%;  Score 221;  DB 2;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    7
ENTRY           S32559     #type complete
TLE             biglycan precursor - bovine
TERNATE_NAMES   dermatan sulfate proteoglycan I (DS-PGI); proteochondroitin
                  core protein; proteoglycan I core protein (PG-I)
ORGANISM        #formal_name Bos primigenius taurus #common_name cattle
DATE            03-May-1994 #sequence_revision 20-Feb-1995 #text_change
                  14-Nov-1997
ACCESSIONS      S32559; S34229; A33701; A31430; PT0078; S55673; A33137
REFERENCE       S32559
   #authors     Torok, M.A.; Evans, S.A.S.; Marcum, J.A.
   #journal     Biochim. Biophys. Acta (1993) 1173:81-84
   #title       cDNA sequence for bovine biglycan (PGI) protein core.
   #accession   S32559
      ##molecule_type mRNA
      ##residues      1-369 ##label TOR
      ##cross-references EMBL:L07953; NID:g162746
      ##experimental_source aortic smooth muscle
REFERENCE       S34229
   #authors     Marcum, J.A.; Torok, M.; Evans, S.
   #submission  submitted to the EMBL Data Library, December 1992
   #accession   S34229
      ##molecule_type mRNA
      ##residues      1-250,'V',252-369 ##label MAR
      ##cross-references EMBL:L07953
REFERENCE       A33701
   #authors     Neame, P.J.; Choi, H.U.; Rosenberg, L.C.
   #journal     J. Biol. Chem. (1989) 264:8653-8661
   #title       The primary structure of the core protein of the small,
                  leucine-rich proteoglycan (PG I) from bovine articular
                  cartilage.
   #cross-references MUID:89255324
   #accession   A33701
      ##molecule_type protein
      ##residues      38-187,'E',189-367,'Y' ##label NEA
      ##experimental_source cartilage
REFERENCE       A31430
   #authors     Choi, H.U.; Johnson, T.L.; Pal, S.; Tang, L.H.; Rosenberg,
                  L.; Neame, P.J.
   #journal     J. Biol. Chem. (1989) 264:2876-2884
   #title       Characterization of the dermatan sulfate proteoglycans,
                  DS-PGI and DS-PGII, from bovine articular cartilage and
                  skin isolated by octyl-sepharose chromatography.
   #cross-references MUID:89123388
   #accession   A31430
      ##molecule_type protein
      ##residues      38-41,'X',43-47,'X',49-63 ##label CHO
      ##note          sequences from skin and cartilage were identical
REFERENCE       PT0077
   #authors     Marcum, J.A.; Thompson, M.A.
   #journal     Biochem. Biophys. Res. Commun. (1991) 175:706-712
   #title       The amino-terminal region of a proteochondroitin core
                  protein, secreted by aortic smooth muscle cells, shares
                  sequence homology with the pre-propeptide region of the
                  biglycan core protein from human bone.
   #cross-references MUID:91207372
   #accession   PT0078
      ##molecule_type protein
      ##residues      17-24,'F',26-30 ##label MA2
      ##experimental_source aortic smooth muscle
REFERENCE       S55673
   #authors     Scott, P.G.; Nakano, T.; Dodd, C.M.
   #journal     Biochim. Biophys. Acta (1995) 1244:121-128
   #title       Small proteoglycans from different regions of the
                  fibrocartilaginous temporomandibular joint disc.
   #cross-references MUID:95284073
   #accession   S55673
      ##molecule_type protein
      ##residues      38-41,'X',43-47,'X',49-53 ##label SCO
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        cartilage; chondroitin sulfate proteoglycan; dermatan
                  sulfate; extracellular matrix; glycoprotein

FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-37            #domain amino-terminal propeptide #status predicted
                      #label PRO\
   38-369           #product biglycan #status predicted #label MAT\
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH\
   42,48            #binding_site dermatan sulfate (Ser) (covalent) #status
```

```
                    experimental\
   181,199          #binding_site dermatan sulfate (Ser) (covalent) #status
                    predicted\
   271,312          #binding_site carbohydrate (Asn) (covalent) #status
                    predicted
SUMMARY          #length 369  #molecular-weight 41590  #checksum 1525

  Query Match           30.3%;  Score 221;  DB 2;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    8
ENTRY           A34548     #type complete
TITLE           follitropin receptor precursor - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
                  29-Jan-1999
ACCESSIONS      A34548; A41729
REFERENCE       A34548
   #authors     Sprengel, R.; Braun, T.; Nikolics, K.; Segaloff, D.L.;
                  Seeburg, P.H.
   #journal     Mol. Endocrinol. (1990) 4:525-530
   #title       The testicular receptor for follicle stimulating hormone:
                  structure and functional expression of cloned cDNA.
   #cross-references MUID:91125358
   #accession   A34548
      ##molecule_type mRNA
      ##residues      1-692 ##label SPR
      ##cross-references GB:L02842; NID:g204183; PID:g204184
REFERENCE       A41729
   #authors     Heckert, L.L.; Daley, I.J.; Griswold, M.D.
   #journal     Mol. Endocrinol. (1992) 6:70-80
   #title       Structural organization of the follicle-stimulating hormone
                  receptor gene.
   #cross-references MUID:92149579
   #accession   A41729
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-692 ##label HEC
      ##cross-references GB:S81198; NID:g245344; PID:g245345
      ##note          sequence inconsistent with the nucleotide translation
      ##note          sequence extracted from NCBI backbone (NCBIN:81117,
                       NCBIN:81119, NCBIN:81121, NCBIN:81171, NCBIN:81174,
                       NCBIN:81178, NCBIN:81183, NCBIN:81185, NCBIN:81194,
                       NCBIN:81198, NCBIP:81116)
REFERENCE       A57562
   #authors     Davis, D.; Liu, X.; Segaloff, D.L.
   #journal     Mol. Endocrinol. (1995) 9:159-170
   #title       Identification of the sites of N-linked glycosylation on the
                  follicle-stimulating hormone (FSH) receptor and assessment
                  of their role in FSH receptor function.
   #contents    annotation; glycosylation sites
FUNCTION
   #description receptor that mediates the biochemical effects of follitropin
CLASSIFICATION  #superfamily glycoprotein hormone receptor; leucine-rich
                  alpha-2-glycoprotein repeat homology
KEYWORDS        alternative splicing; G protein-coupled receptor;
                  glycoprotein; hormone receptor; phosphoprotein;
                  transmembrane protein
FEATURE
   1-15             #domain signal sequence #status predicted #label SIG\
   16-692           #product follitropin receptor #status predicted #label
                      MAT\
   16-366           #domain extracellular hormone binding #status predicted
                      #label EHB\
   56-70            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   71-95            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   96-120           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   121-145          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   146-169          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   172-193          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   194-218          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   367-387          #domain transmembrane #status predicted #label TM1\
   398-421          #domain transmembrane #status predicted #label TM2\
   443-465          #domain transmembrane #status predicted #label TM3\
   486-508          #domain transmembrane #status predicted #label TM4\
   529-550          #domain transmembrane #status predicted #label TM5\
   574-597          #domain transmembrane #status predicted #label TM6\
   609-630          #domain transmembrane #status predicted #label TM7\
   191,199,293      #binding_site carbohydrate (Asn) (covalent) #status
                      predicted\
   554              #binding_site phosphate (Thr) (covalent) (by protein
                      kinase C) #status predicted\
   595              #binding_site phosphate (Ser) (covalent) (by protein
                      kinase C) #status predicted
SUMMARY          #length 692  #molecular-weight 77680  #checksum 8898

  Query Match           30.3%;  Score 221;  DB 2;  Length 692;
  Best Local Similarity  32.4%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQDSKVTEIPTDLPRNAIELRFVLTKLRVIPKGSFAGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   :|:|:    :: || |:|: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPSLRYLLI 127
          : | :| |  |  |  : :  :|::  |    |: | ||: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT    9
ENTRY           JC2237     #type complete
TITLE           follitropin receptor, testis - horse
ALTERNATE_NAMES eFSHR
ORGANISM        #formal_name Equus caballus #common_name domestic horse
DATE            28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change
                  29-Jan-1999
ACCESSIONS      JC2237; JC2370
REFERENCE       JC2237
   #authors     Robert, P.; Amsellem, S.; Christophe, S.; Benifla, J.L.;
                  Bellet, D.; Koman, A.; Bidart, J.M.
   #journal     Biochem. Biophys. Res. Commun. (1994) 201:201-207
   #title       Cloning and sequencing of the equine testicular follitropin
                  receptor.
   #cross-references MUID:94256980
   #accession   JC2237
      ##molecule_type mRNA
      ##residues      1-694 ##label ROB
      ##cross-references GB:S70150; NID:g546896; PID:g546897
      ##experimental_source testis
CLASSIFICATION  #superfamily glycoprotein hormone receptor; leucine-rich
                  alpha-2-glycoprotein repeat homology
KEYWORDS        glycoprotein; hormone receptor; transmembrane protein
FEATURE
   56-70            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   71-95            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   96-120           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
```

```
   121-145          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   146-169          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   172-193          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   194-218          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   366-386          #domain transmembrane #status predicted #label TM1\
   398-420          #domain transmembrane #status predicted #label TM2\
   443-464          #domain transmembrane #status predicted #label TM3\
   485-507          #domain transmembrane #status predicted #label TM4\
   528-549          #domain transmembrane #status predicted #label TM5\
   573-596          #domain transmembrane #status predicted #label TM6\
   608-629          #domain transmembrane #status predicted #label TM7\
   191,199,268,293  #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY             #length 694  #molecular-weight 78004  #checksum 8235

  Query Match           30.3%;  Score 221;  DB 2;  Length 694;
  Best Local Similarity  32.4%;  Pred. No. 6.69e-18;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNALELRFVLTKLRVIPKGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   :|:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDHDAFQNLPNLQYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|| ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT  10
ENTRY               I45896     #type complete
TITLE               follicle stimulating hormone receptor - bovine
ORGANISM            #formal_name Bos primigenius taurus #common_name cattle
DATE                15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change
                      06-Jun-1997
ACCESSIONS          I45896
REFERENCE           I45896
  #authors          Houde, A.; Lambert, A.; Saumande, J.; Silversides, D.W.;
                      Lussier, J.G.
  #journal          Mol. Reprod. Dev. (1994) 39:127-135
  #title            Structure of the bovine follicle-stimulating hormone receptor
                      complementary DNA and expression in bovine tissues.
  #cross-references MUID:95127199
  #accession        I45896
    ##status         preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues       1-695 ##label HOU
    ##cross-references GB:L22319; NID:g404671; PID:g404672
GENETICS
  #gene             FSHR
CLASSIFICATION      #superfamily glycoprotein hormone receptor; leucine-rich
                      alpha-2-glycoprotein repeat homology
SUMMARY             #length 695  #molecular-weight 78084  #checksum 895

  Query Match           30.1%;  Score 220;  DB 2;  Length 695;
  Best Local Similarity  32.4%;  Pred. No. 9.58e-18;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQESKVTEIPSDLPRDAVELRFVLTKLRVIPKGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT  11
ENTRY               JC1493     #type complete
TITLE               follitropin receptor - sheep
ALTERNATE_NAMES     follicle stimulating hormone receptor
ORGANISM            #formal_name Ovis orientalis aries, Ovis ammon aries
                      #common_name domestic sheep
DATE                03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                      17-Mar-1999
ACCESSIONS          JC1493; I47080
REFERENCE           JC1493
  #authors          Khan, H.; Yarney, T.A.; Sairam, M.R.
  #journal          Biochem. Biophys. Res. Commun. (1993) 190:888-894
  #title            Cloning of alternately spliced mRNA transcripts coding for
                      variants of ovine testicular follitropin receptor lacking
                      the G protein coupling domains.
  #cross-references MUID:93176195
  #accession        JC1493
    ##molecule_type mRNA
    ##residues       1-695 ##label KHA
    ##experimental_source testis
REFERENCE           I47080
  #authors          Yarney, T.A.; Sairam, M.R.; Khan, H.; Ravindranath, N.;
                      Payne, S.; Seidah, N.G.
  #journal          Mol. Cell. Endocrinol. (1993) 93:219-226
  #title            Molecular cloning and expression of the ovine testicular
                      follicle stimulating hormone receptor.
  #cross-references MUID:93351750
  #accession        I47080
    ##status         preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues       1-695 ##label YAR
    ##cross-references GB:L07302; NID:g165884; PID:g165885
GENETICS
  #gene             FSH-R
CLASSIFICATION      #superfamily glycoprotein hormone receptor; leucine-rich
                      alpha-2-glycoprotein repeat homology
KEYWORDS            G protein-coupled receptor; glycoprotein; transmembrane
                      protein
FEATURE
  191,199             #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY             #length 695  #molecular-weight 78237  #checksum 1112

  Query Match           30.0%;  Score 219;  DB 2;  Length 695;
  Best Local Similarity  31.4%;  Pred. No. 1.37e-17;
  Matches   33;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQDSKVTEMPSDLPRDAVELRFVLTKLRVIPEGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |:|::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT  12
ENTRY               JN0898     #type complete
TITLE               follitropin receptor precursor - crab-eating macaque
ALTERNATE_NAMES     follicle-stimulating hormone receptor (FSHR)
ORGANISM            #formal_name Macaca fascicularis #common_name crab-eating
                      macaque
DATE                19-May-1994 #sequence_revision 19-May-1994 #text_change
                      13-Nov-1998
ACCESSIONS          JN0898; S36452
REFERENCE           JN0898
  #authors          Gromoll, J.; Dankbar, B.; Sharma, R.S.; Nieschlag, E.
  #journal          Biochem. Biophys. Res. Commun. (1993) 196:1066-1072
  #title            Molecular cloning of the testicular follicle stimulating
                      hormone receptor of the non human primate Macaca
                      fascicularis and identification of multiple transcripts in
                      the testis.
  #cross-references MUID:94071854
  #accession        JN0898
    ##molecule_type mRNA
    ##residues       1-695 ##label GRO
```

```
     ##cross-references EMBL:X74454; NID:g396801; PID:g396802
     ##note          the authors translated the codon AGT for residue 488 as
                      Arg
FUNCTION
   #description  receptor that mediates the biochemical effects of follitropin
CLASSIFICATION   #superfamily glycoprotein hormone receptor; leucine-rich
                  alpha-2-glycoprotein repeat homology
KEYWORDS         G protein-coupled receptor; glycoprotein; hormone receptor;
                  phosphoprotein; pituitary; transmembrane protein
FEATURE
   1-17             #domain signal sequence #status predicted #label SIG\
   18-695           #product follitropin receptor #status predicted #label
                      PFH\
   71-95            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   96-120           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   121-145          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   146-169          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   172-193          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   194-218          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   367-387          #domain transmembrane #status predicted #label TM1\
   399-421          #domain transmembrane #status predicted #label TM2\
   444-465          #domain transmembrane #status predicted #label TM3\
   486-508          #domain transmembrane #status predicted #label TM4\
   529-550          #domain transmembrane #status predicted #label TM5\
   574-597          #domain transmembrane #status predicted #label TM6\
   609-630          #domain transmembrane #status predicted #label TM7\
   191,199,293,318  #binding_site carbohydrate (Asn) (covalent) #status
                      predicted\
   555              #binding_site phosphate (Thr) (covalent) (by protein
                      kinase C) #status predicted\
   596              #binding_site phosphate (Ser) (covalent) (by protein
                      kinase C) #status predicted
SUMMARY          #length 695  #molecular-weight 78343  #checksum 1011

  Query Match           29.9%;  Score 218;  DB 2;  Length 695;
  Best Local Similarity  31.4%;  Pred. No. 1.96e-17;
  Matches   33;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNAIELRFVHTKLRVIQKGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   :|:|: :  :: |  |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPNLRYLLI 127
          : | :| |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   13
ENTRY            QRHUFT     #type complete
TITLE            follitropin receptor precursor - human
ALTERNATE_NAMES  follicle stimulating hormone receptor (FSHR)
CONTAINS         follitropin receptor precursor long splice form; follitropin
                   receptor precursor short splice form
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             30-Sep-1991 #sequence_revision 06-Sep-1996 #text_change
                   05-Sep-1997
ACCESSIONS       I57661; I56448; PC1147; S30560; I57672; JN0122
REFERENCE        I57661
   #authors      Gromoll, J.; Dankbar, B.; Gudermann, T.
   #journal      Mol. Cell. Endocrinol. (1994) 102:93-102
   #title        Characterization of the 5' flanking region of the human
                   follicle-stimulating hormone receptor gene.
   #cross-references MUID:95011044
   #accession    I57661
      ##status       translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues     1-51 ##label GRO
      ##cross-references GB:S73199; NID:g685036; PID:g685037
REFERENCE        I56448
   #authors      Gromoll, J.; Ried, T.; Holtgreve-Grez, H.; Nieschlag, E.;
                   Gudermann, T.
   #journal      J. Mol. Endocrinol. (1994) 12:265-271
   #title        Localization of the human FSH receptor to chromosome 2 p21
                   using a genomic probe comprising exon 10.
   #cross-references MUID:95000244
   #accession    I56448
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues     286-695 ##label GR2
      ##cross-references GB:S73526; NID:g688069; PID:g688070
REFERENCE        PC1147
   #authors      Gromoll, J.; Gudermann, T.; Nieschlag, E.
   #journal      Biochem. Biophys. Res. Commun. (1992) 188:1077-1083
   #title        Molecular cloning of a truncated isoform of the human
                   follicle stimulating hormone receptor.
   #cross-references MUID:93075197
   #accession    PC1147
      ##status       nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues     1-223,286-294,'P',296-342 ##label GR3
      ##cross-references EMBL:X68044; NID:g31473; PID:g31474
      ##experimental_source testis
REFERENCE        S30560
   #authors      Gromoll, J.
   #submission   submitted to the EMBL Data Library, August 1992
   #accession    S30560
      ##molecule_type mRNA
      ##residues     1-12,'R',14-223,286-294,'P',296-342 ##label GR4
      ##cross-references EMBL:X68044; NID:g31473; PID:g31474
REFERENCE        I57672
   #authors      Kelton, C.A.; Cheng, S.V.; Nugent, N.P.; Schweickhardt, R.L.;
                   Rosenthal, J.L.; Overton, S.A.; Wands, G.D.; Kuzeja, J.B.;
                   Luchette, C.A.; Chappel, S.C.
   #journal      Mol. Cell. Endocrinol. (1992) 89:141-151
   #title        The cloning of the human follicle stimulating hormone
                   receptor and its expression in COS-7, CHO, and Y-1 cells.
   #cross-references MUID:93246012
   #accession    I57672
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-679,'N',681-695 ##label KEL
      ##cross-references GB:S59900; NID:g300072; PID:g300073
REFERENCE        JN0122
   #authors      Minegish, T.; Nakamura, K.; Takakura, Y.; Ibuki, Y.;
                   Igarashi, M.
   #journal      Biochem. Biophys. Res. Commun. (1991) 175:1125-1130
   #title        Cloning and sequencing of human FSH receptor cDNA.
   #cross-references MUID:91222171
   #accession    JN0122
      ##molecule_type mRNA
      ##residues     1-111,'T',113-196,'AV',199-306,'A',308-695 ##label MIN
      ##cross-references EMBL:M65085; NID:g182770; PID:g182771
GENETICS
   #gene         GDB:FSHR
      ##cross-references GDB:127510; OMIM:136435
   #map_position 2p21-2p16
   #introns      223/3
   #note         the exact position of the intron cannot be determined from
                   the experimental data
FUNCTION
   #description  receptor that mediates the biochemical effects of follitropin
CLASSIFICATION   #superfamily glycoprotein hormone receptor; leucine-rich
                   alpha-2-glycoprotein repeat homology
KEYWORDS         alternative splicing; G protein-coupled receptor;
                   glycoprotein; hormone receptor; phosphoprotein;
                   transmembrane protein
FEATURE
   1-695            #product follitropin receptor precursor, long splice
                      form #status predicted #label SPL1\
   1-223,286-695    #product follitropin receptor precursor, short splice
```

```
                  form #status predicted #label SPL2\
   1-15           #domain signal sequence #status predicted #label SIG\
   16-695         #product follitropin receptor #status predicted #label
                    MAT\
   16-366         #domain extracellular hormone binding #status predicted
                    #label EHB\
   56-70          #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR1\
   71-95          #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR2\
   96-120         #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR3\
   121-145        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR4\
   146-169        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR5\
   172-193        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR6\
   194-218        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR7\
   367-387        #domain transmembrane #status predicted #label TM1\
   398-421        #domain transmembrane #status predicted #label TM2\
   444-465        #domain transmembrane #status predicted #label TM3\
   486-508        #domain transmembrane #status predicted #label TM4\
   529-550        #domain transmembrane #status predicted #label TM5\
   574-597        #domain transmembrane #status predicted #label TM6\
   609-630        #domain transmembrane #status predicted #label TM7\
   191,199,293,318 #binding_site carbohydrate (Asn) (covalent) #status
                    predicted\
   555            #binding_site phosphate (Thr) (covalent) (by protein
                    kinase C) #status predicted\
   596            #binding_site phosphate (Ser) (covalent) (by protein
                    kinase C) #status predicted
SUMMARY         #length 695  #molecular-weight 78267  #checksum 2066

  Query Match           29.7%;  Score 217;  DB 1;  Length 695;
  Best Local Similarity  32.4%;  Pred. No. 2.80e-17;
  Matches    34;  Conservative   26;  Mismatches  43;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNAIELRFVLTKLRVIQKGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   :|:|:    :: |  |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPNLQYLLI 127
          : | :| |  |  |  : :  :|::  |    |: | :|| ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   14
ENTRY           A58532     #type complete
TITLE           glial cell membrane glycoprotein LIG-1 precursor - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change
                  17-Mar-1999
ACCESSIONS      A58532
REFERENCE       A58532
   #authors     Suzuki, Y.; Sato, N.; Tohyama, M.; Wanaka, A.; Takagi, T.
   #journal     J. Biol. Chem. (1996) 271:22522-22527
   #title       cDNA cloning of a novel membrane glycoprotein that is
                  expressed specifically in glial cells in the mouse brain;
                  LIG-1, a protein with leucine-rich repeats and
                  immunoglobulin-like domains.
   #cross-references MUID:96394313
   #accession   A58532
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-1091 ##label SUZ
      ##cross-references GB:D78572; NID:g1545806; PID:g1545807
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
FEATURE
   36-61           #domain proteoglycan amino-terminal homology #label PAH\
   71-94           #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR1\
   95-117          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR2\
   118-141         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR3\
   142-165         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR4\
   166-189         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR5\
   191-213         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR6\
   214-237         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR7\
   238-261         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR8\
   262-285         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR9\
   286-309         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR10\
   310-333         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR11\
   334-357         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR12\
   358-381         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR13\
   385-408         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR14\
   409-432         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR15\
   440-485         #domain proteoglycan carboxyl-terminal homology #label
                     PCS1
SUMMARY         #length 1091  #molecular-weight 119283  #checksum 7937

  Query Match           29.3%;  Score 214;  DB 2;  Length 1091;
  Best Local Similarity  42.5%;  Pred. No. 8.16e-17;
  Matches    37;  Conservative   19;  Mismatches  27;  Indels   4;  Gaps   3;

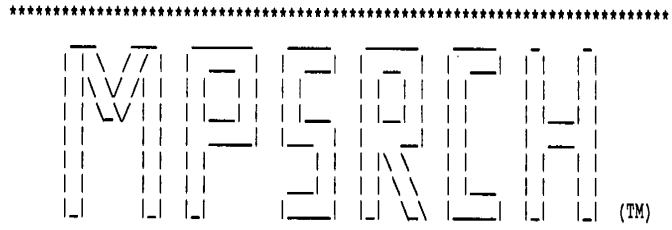
Db    330 AELS-SLSILRLSHNAISHIAEGAFKGLKSLRVLDLDHNEISGTIEDTSGAFTGLDNLSK 388
          |:|: ::  :|| :|:|  |: |||   | |: :|:  |:|| |  :: || || :|:
Qy     22 ANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQISD-I--APDAFQGLKSLTS 78

Db    389 LTLFGNKIKSVAKRAFSGLESLEHLNL 415
          | |:||||  :||  | || ||: | |
Qy     79 LVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT   15
ENTRY           G02020     #type complete
TITLE           p37NB - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
                  16-Dec-1998
ACCESSIONS      G02020
REFERENCE       H00740
   #authors     Kim, D.; LaQuaglia, M.; Yang, S.Y.
   #submission  submitted to the EMBL Data Library, August 1995
   #accession   G02020
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-313 ##label KIM
      ##cross-references EMBL:U32907; NID:g1236328; PID:g1236329
CLASSIFICATION  #superfamily proteoglycan carboxyl-terminal homology
FEATURE
   161-207         #domain proteoglycan carboxyl-terminal homology #label
                     PCS1
SUMMARY         #length 313  #molecular-weight 36287  #checksum 5976

  Query Match           28.6%;  Score 209;  DB 2;  Length 313;
  Best Local Similarity  32.9%;  Pred. No. 4.81e-16;
  Matches    28;  Conservative   27;  Mismatches  30;  Indels   0;  Gaps   0;

Db     65 LDCQERKLVYVLPGWPQDLLHMLLARNKIRTLKNNMFSKFKKLKSLDLQQNEISKIESEA 124
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:43:11 1999; MasPar time 5.02 Seconds
591.610 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-3
Description: (1-105) from US09191647.pep
Perfect Score: 730
Sequence: 1 SPCTCSNNIVDCRGKGLMEI.....ITEIAKGLFDGLVSLQLLLL 105

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 43.647; Variance 83.058; scale 0.525

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	459	62.9	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	2.99e-69
2	221	30.3	368	1	PGS1_HUMAN	BONE/CARTILAGE PROTEOG	7.55e-22
3	221	30.3	369	1	PGS1_MOUSE	BONE/CARTILAGE PROTEOG	7.55e-22
4	221	30.3	369	1	PGS1_RAT	BONE/CARTILAGE PROTEOG	7.55e-22
5	221	30.3	369	1	PGS1_CANFA	BONE/CARTILAGE PROTEOG	7.55e-22
6	221	30.3	692	1	FSHR_RAT	FOLLICLE STIMULATING H	7.55e-22
7	221	30.3	694	1	FSHR_HORSE	FOLLICLE STIMULATING H	7.55e-22
8	220	30.1	695	1	FSHR_BOVIN	FOLLICLE STIMULATING H	1.16e-21
9	219	30.0	695	1	FSHR_SHEEP	FOLLICLE STIMULATING H	1.77e-21
10	219	30.0	695	1	FSHR_PIG	FOLLICLE STIMULATING H	1.77e-21
11	218	29.9	695	1	FSHR_MACFA	FOLLICLE STIMULATING H	2.72e-21
12	217	29.7	695	1	FSHR_HUMAN	FOLLICLE STIMULATING H	4.16e-21
13	215	29.5	687	1	FSHR_EQUAS	FOLLICLE STIMULATING H	9.72e-21
14	211	28.9	369	1	PGS1_BOVIN	BONE/CARTILAGE PROTEOG	5.29e-20
15	204	27.9	357	1	PGS2_CHICK	BONE PROTEOGLYCAN II P	1.01e-18
16	200	27.4	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	5.37e-18
17	200	27.4	662	1	GARP_HUMAN	GARP PROTEIN PRECURSOR	5.37e-18
18	198	27.1	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	1.24e-17
19	197	27.0	354	1	PGS2_RAT	BONE PROTEOGLYCAN II P	1.87e-17
20	193	26.4	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	9.82e-17
21	192	26.3	693	1	FSHR_CHICK	FOLLICLE STIMULATING H	1.48e-16
22	190	26.0	560	1	GPV_HUMAN	PLATELET GLYCOPROTEIN	3.38e-16
23	189	25.9	360	1	PGS2_CANFA	BONE PROTEOGLYCAN II P	5.10e-16
24	189	25.9	567	1	GPV_RAT	PLATELET GLYCOPROTEIN	5.10e-16
25	188	25.8	359	1	PGS2_HUMAN	BONE PROTEOGLYCAN II P	7.70e-16
26	187	25.6	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	1.16e-15
27	186	25.5	354	1	PGS2_MOUSE	BONE PROTEOGLYCAN II P	1.75e-15
28	184	25.2	361	1	CHAD_BOVIN	CHONDROADHERIN PRECURS	3.95e-15
29	182	24.9	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	8.92e-15
30	179	24.5	382	1	PARG_HUMAN	PROLARGIN PRECURSOR (P	3.01e-14
31	175	24.0	1115	1	GPCR_LYMST	G-PROTEIN COUPLED RECE	1.51e-13
32	170	23.3	536	1	CBP8_HUMAN	CARBOXYPEPTIDASE N 83	1.12e-12
33	170	23.3	1134	1	CHAO_DROME	CHAOPTIN PRECURSOR (PH	1.12e-12
34	168	23.0	567	1	GPV_MOUSE	PLATELET GLYCOPROTEIN	2.48e-12
35	155	21.2	682	1	CONN_DROME	CONNECTIN PRECURSOR.	4.03e-10
36	152	20.8	626	1	GPBA_HUMAN	PLATELET GLYCOPROTEIN	1.28e-09
37	148	20.3	343	1	LUM_CHICK	LUMICAN PRECURSOR (LUM	5.88e-09
38	148	20.3	375	1	FMOD_BOVIN	FIBROMODULIN PRECURSOR	5.88e-09
39	148	20.3	440	1	OMGP_HUMAN	OLIGODENDROCYTE-MYELIN	5.88e-09
40	148	20.3	440	1	OMGP_MOUSE	OLIGODENDROCYTE-MYELIN	5.88e-09
41	142	19.5	177	1	GPIX_HUMAN	PLATELET GLYCOPROTEIN	5.64e-08
42	141	19.3	376	1	FMOD_RAT	FIBROMODULIN PRECURSOR	8.20e-08
43	141	19.3	376	1	FMOD_MOUSE	FIBROMODULIN PRECURSOR	8.20e-08
44	137	18.8	376	1	FMOD_HUMAN	FIBROMODULIN PRECURSOR	3.61e-07
45	137	18.8	925	1	GLHR_ANTEL	PROBABLE GLYCOPROTEIN	3.61e-07

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          62.9%;  Score 459;  DB 1;  Length 1480;
  Best Local Similarity  53.8%;  Pred. No. 2.99e-69;
  Matches   56;  Conservative   24;  Mismatches  24;  Indels   0;  Gaps   0;

Db    298 PCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRIDLSNNN 357
          || |:: ||||| |:|  :|  ||:   ::||||| |  :|: :|: :::|:|||:|:|
Qy      2 PCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQ 61

Db    358 ISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLL 401
          || || ||: ||| ||:|||||||| ::: |:| || ||:||||
Qy     62 ISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    2
ID   PGS1_HUMAN     STANDARD;      PRT;   368 AA.
AC   P21810; P13247;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE;
RX   MEDLINE; 89174714.
RA   FISHER L.W., TERMINE J.D., YOUNG M.F.;
RT   "Deduced protein sequence of bone small proteoglycan I (biglycan)
RT   shows homology with proteoglycan II (decorin) and several
RT   nonconnective tissue proteins in a variety of species.";
RL   J. BIOL. CHEM. 264:4571-4576(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91317791.
RA   FISHER L.W., HEEGAARD A.M., VETTER U., VOGEL W., JUST W.,
RA   TERMINE J.D., YOUNG M.F.;
RT   "Human biglycan gene. Putative promoter, intron-exon junctions, and
RT   chromosomal localization.";
RL   J. BIOL. CHEM. 266:14371-14377(1991).
RN   [3]
RP   SEQUENCE OF 38-57.
RX   MEDLINE; 90073579.
RA   ROUGHLEY P.J., WHITE R.J.;
RT   "Dermatan sulphate proteoglycans of human articular cartilage. The
RT   properties of dermatan sulphate proteoglycans I and II.";
RL   BIOCHEM. J. 262:823-827(1989).
RN   [4]
RP   SEQUENCE OF 38-66.
```

```
RX   MEDLINE; 87250639.
RA   FISHER L.W., HAWKINS G.R., TUROSS N., TERMINE J.D.;
RT   "Purification and partial characterization of small proteoglycans I
RT   and II, bone sialoproteins I and II, and osteonectin from the mineral
RT   compartment of developing human bone.";
RL   J. BIOL. CHEM. 262:9702-9708(1987).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04599; G306884; -.
DR   EMBL; M65153; G179433; ALT_SEQ.
DR   EMBL; M65152; G179433; JOINED.
DR   PIR; A28457; A28457.
DR   PIR; A32458; A32458.
DR   PIR; A40757; A40757.
DR   PIR; S05639; S05639.
DR   MIM; 301870; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    19       POTENTIAL.
FT   PROPEP      20    37
FT   CHAIN       38   368       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      71   342       LEUCINE-RICH REPEATS.
FT   REPEAT      71    94       LRR 1.
FT   REPEAT      95   115       LRR 2.
FT   REPEAT     116   139       LRR 3.
FT   REPEAT     140   163       LRR 4.
FT   REPEAT     164   184       LRR 5.
FT   REPEAT     185   210       LRR 6.
FT   REPEAT     211   230       LRR 7.
FT   REPEAT     231   254       LRR 8.
FT   REPEAT     255   275       LRR 9.
FT   REPEAT     276   301       LRR 10.
FT   REPEAT     302   320       LRR 11.
FT   REPEAT     335   342       LRR 12.
FT   CARBOHYD    42    42       GLYCOSAMINOGLYCAN.
FT   CARBOHYD    47    47       GLYCOSAMINOGLYCAN.
FT   CARBOHYD   270   270       POTENTIAL.
FT   CARBOHYD   311   311       POTENTIAL.
FT   DISULFID    63    76       BY SIMILARITY.
FT   DISULFID   321   354       BY SIMILARITY.
FT   CONFLICT   139   140       KL -> NV (IN REF. 1).
FT   CONFLICT   163   164       EL -> DV (IN REF. 1).
SQ   SEQUENCE   368 AA;  41654 MW;  6820F8DF CRC32;

  Query Match          30.3%;  Score 221;  DB 1;  Length 368;
  Best Local Similarity  38.6%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    150 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 209
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    210 -LNYLRISEAKLTGIPKDLPETLNELHL 236
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102
```

```
RESULT   3
ID   PGS1_MOUSE     STANDARD;      PRT;   369 AA.
AC   P28653; Q61355;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS; TISSUE=FIBROBLAST;
RA   NAITOH Y., SUZUKI S.;
RL   SUBMITTED (JUL-1990) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS; TISSUE=EMBRYO;
RX   MEDLINE; 94319093.
RA   RAU W., JUST W., VETTER U., VOGEL W.;
RT   "A dinucleotide repeat in the mouse biglycan gene (EST) on the X
RT   chromosome.";
RL   MAMM. GENOME 5:395-396(1994).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53928; G53667; -.
DR   EMBL; L20276; G348962; -.
DR   PIR; S20811; S20811.
DR   MGD; MGI:88158; BGN.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    19       POTENTIAL.
FT   PROPEP      20    37
FT   CHAIN       38   369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72   343       LEUCINE-RICH REPEATS.
FT   REPEAT      72    95       LRR 1.
FT   REPEAT      96   116       LRR 2.
FT   REPEAT     117   140       LRR 3.
FT   REPEAT     141   164       LRR 4.
FT   REPEAT     165   185       LRR 5.
FT   REPEAT     186   211       LRR 6.
FT   REPEAT     212   231       LRR 7.
FT   REPEAT     232   255       LRR 8.
FT   REPEAT     256   276       LRR 9.
FT   REPEAT     277   302       LRR 10.
FT   REPEAT     303   321       LRR 11.
FT   REPEAT     336   343       LRR 12.
FT   CARBOHYD    42    42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48    48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271   271       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   DISULFID    64    77       BY SIMILARITY.
FT   DISULFID   322   355       BY SIMILARITY.
FT   CONFLICT    68    68       C -> W (IN REF. 2).
SQ   SEQUENCE   369 AA;  41639 MW;  ED21DD6B CRC32;
```

```
  Query Match          30.3%;  Score 221;  DB 1;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   4
ID   PGS1_RAT       STANDARD;      PRT;   369 AA.
AC   P47853;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=VASCULAR SMOOTH MUSCLE;
RX   MEDLINE; 91184222.
RA   DREHER K.L., ASUNDI V., MATZURA D., COWAN K.;
RT   "Vascular smooth muscle biglycan represents a highly conserved
RT   proteoglycan within the arterial wall.";
RL   EUR. J. CELL BIOL. 53:296-304(1990).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17834; G600498; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     19       POTENTIAL.
FT   PROPEP      20     37
FT   CHAIN       38    369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72    343       LEUCINE-RICH REPEATS.
FT   REPEAT      72     95       LRR 1.
FT   REPEAT      96    116       LRR 2.
FT   REPEAT     117    140       LRR 3.
FT   REPEAT     141    164       LRR 4.
FT   REPEAT     165    185       LRR 5.
FT   REPEAT     186    211       LRR 6.
FT   REPEAT     212    231       LRR 7.
FT   REPEAT     232    255       LRR 8.
FT   REPEAT     256    276       LRR 9.
FT   REPEAT     277    302       LRR 10.
FT   REPEAT     303    321       LRR 11.
FT   REPEAT     336    343       LRR 12.
FT   CARBOHYD    42     42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48     48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271    271       POTENTIAL.
FT   CARBOHYD   312    312       POTENTIAL.
FT   DISULFID    64     77       BY SIMILARITY.
FT   DISULFID   322    355       BY SIMILARITY.
SQ   SEQUENCE   369 AA;  41706 MW;  6555ECED CRC32;

  Query Match          30.3%;  Score 221;  DB 1;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           |  | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   5
ID   PGS1_CANFA     STANDARD;      PRT;   369 AA.
AC   O02678;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   CANIS FAMILIARIS (DOG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GLANT T.T.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U83140; G1916846; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     19       POTENTIAL.
FT   PROPEP      20     37       BY SIMILARITY.
FT   CHAIN       38    369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72    343       LEUCINE-RICH REPEATS.
FT   REPEAT      72     95       LRR 1.
FT   REPEAT      96    116       LRR 2.
FT   REPEAT     117    140       LRR 3.
FT   REPEAT     141    164       LRR 4.
FT   REPEAT     165    185       LRR 5.
FT   REPEAT     186    211       LRR 6.
FT   REPEAT     212    231       LRR 7.
FT   REPEAT     232    255       LRR 8.
FT   REPEAT     256    276       LRR 9.
FT   REPEAT     277    302       LRR 10.
FT   REPEAT     303    321       LRR 11.
FT   REPEAT     336    343       LRR 12.
FT   CARBOHYD    42     42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48     48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271    271       POTENTIAL.
FT   CARBOHYD   312    312       POTENTIAL.
```

```
FT   DISULFID    64     77       BY SIMILARITY.
FT   DISULFID    322    355      BY SIMILARITY.
SQ   SEQUENCE   369 AA;  41566 MW;  F794CEEA CRC32;

  Query Match          30.3%;  Score 221;  DB 1;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
        |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
          | | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   6
ID   FSHR_RAT       STANDARD;      PRT;   692 AA.
AC   P20395;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SERTOLI CELLS;
RX   MEDLINE; 91125358.
RA   SPRENGEL R., BRAUN T., NIKOLICS K., SEGALOFF D.L., SEEBURG P.H.;
RT   "The testicular receptor for follicle stimulating hormone: structure
RT   and functional expression of cloned cDNA.";
RL   MOL. ENDOCRINOL. 4:525-530(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92149579.
RA   HECKERT L.L., DALEY I.J., GRISWOLD M.D.;
RT   "Structural organization of the follicle-stimulating hormone receptor
RT   gene.";
RL   MOL. ENDOCRINOL. 6:70-80(1992).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SERTOLI CELLS AND OVARIAN GRANULOSA CELLS.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L02842; G204184; -.
DR   PIR; A34548; A34548.
DR   PIR; A41729; A41729.
DR   GCRDB; GCR_0234; -.
DR   GCRDB; GCR_0456; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    692       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18    365       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    366   386       1 (POTENTIAL).
FT   DOMAIN      387   397       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    398   420       2 (POTENTIAL).
FT   DOMAIN      421   442       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    443   464       3 (POTENTIAL).
FT   DOMAIN      465   484       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    485   507       4 (POTENTIAL).
FT   DOMAIN      508   527       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    528   549       5 (POTENTIAL).
FT   DOMAIN      550   572       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    573   596       6 (POTENTIAL).
FT   DOMAIN      597   607       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    608   629       7 (POTENTIAL).
FT   DOMAIN      630   692       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    191   191       POTENTIAL.
FT   CARBOHYD    199   199       POTENTIAL.
FT   CARBOHYD    293   293       POTENTIAL.
FT   DISULFID    441   516       BY SIMILARITY.
SQ   SEQUENCE   692 AA;  77681 MW;  BFD1CAD7 CRC32;

  Query Match          30.3%;  Score 221;  DB 1;  Length 692;
  Best Local Similarity  32.4%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQDSKVTEIPTDLPRNAIELRFVLTKLRVIPKGSFAGFGDLEKIEISQNDV 82
          | ||| :  |:   : |||::||   :|:|:    :: || |:|: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPSLRYLLI 127
             : | :| |  |  |  : :  :|::  |    |: | ||: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   7
ID   FSHR_HORSE     STANDARD;      PRT;   694 AA.
AC   P47799;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   EQUUS CABALLUS (HORSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PERISSODACTYLA; EQUIDAE; EQUUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 94256980.
RA   ROBERT P., AMSELLEM S., CHRISTOPHE S., BENIFLA J.L., BELLET D.,
RA   KOMAN A., BIDART J.M.;
RT   "Cloning and sequencing of the equine testicular follitropin
RT   receptor.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 201:201-207(1994).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE. AMONG ALL MAMMALIAN FSH RECEPTORS, ON THE HORSE
CC       RECEPTOR DOES NOT BIND LH/CHORIONIC GONADOTROPHIN (CG).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S70150; G546897; -.
DR   GCRDB; GCR_1251; -.
```

```
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   694       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   365       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   366   386       1 (POTENTIAL).
FT   DOMAIN     387   397       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   398   420       2 (POTENTIAL).
FT   DOMAIN     421   442       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   443   464       3 (POTENTIAL).
FT   DOMAIN     465   484       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   485   507       4 (POTENTIAL).
FT   DOMAIN     508   527       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   528   549       5 (POTENTIAL).
FT   DOMAIN     550   572       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   573   596       6 (POTENTIAL).
FT   DOMAIN     597   607       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   608   629       7 (POTENTIAL).
FT   DOMAIN     630   694       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   191   191       POTENTIAL.
FT   CARBOHYD   199   199       POTENTIAL.
FT   CARBOHYD   268   268       POTENTIAL.
FT   CARBOHYD   293   293       POTENTIAL.
SQ   SEQUENCE   694 AA;  78004 MW;  C20D75BE CRC32;

  Query Match           30.3%;  Score 221;  DB 1;  Length 694;
  Best Local Similarity  32.4%;  Pred. No. 7.55e-22;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNALELRFVLTKLRVIPKGAFSGFGDLEKIEISQNDV 82
           | ||| :  |:   : |||::||   :|:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDHDAFQNLPNLQYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|| ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT    8
ID   FSHR_BOVIN     STANDARD;      PRT;   695 AA.
AC   P35376;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HOLSTEIN; TISSUE=OVARY, AND TESTIS;
RX   MEDLINE; 95127199.
RA   HOUDE A., LAMBERT A., SAUMANDE J., SILVERSIDES D.W., LUSSIER J.G.;
RT   "Structure of the bovine follicle-stimulating hormone receptor
RT   complementary DNA and expression in bovine tissues.";
RL   MOL. REPROD. DEV. 39:127-135(1994).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L22319; G404672; -.
DR   GCRDB; GCR_0766; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367   387       1 (POTENTIAL).
FT   DOMAIN     388   398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399   421       2 (POTENTIAL).
FT   DOMAIN     422   443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444   465       3 (POTENTIAL).
FT   DOMAIN     466   485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486   508       4 (POTENTIAL).
FT   DOMAIN     509   528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529   550       5 (POTENTIAL).
FT   DOMAIN     551   573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574   597       6 (POTENTIAL).
FT   DOMAIN     598   608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609   630       7 (POTENTIAL).
FT   DOMAIN     631   695       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   191   191       POTENTIAL.
FT   CARBOHYD   199   199       POTENTIAL.
FT   CARBOHYD   293   293       POTENTIAL.
SQ   SEQUENCE   695 AA;  78084 MW;  93AEC243 CRC32;

  Query Match           30.1%;  Score 220;  DB 1;  Length 695;
  Best Local Similarity  32.4%;  Pred. No. 1.16e-21;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQESKVTEIPSDLPRDAVELRFVLTKLRVIPKGAFSGFGDLEKIEISQNDV 82
           | ||| :  |:   : |||::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT    9
ID   FSHR_SHEEP     STANDARD;      PRT;   695 AA.
AC   P35379;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; CAPRINAE; OVIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 93351750.
RA   YARNEY T.A., SAIRAM M.R., KHAN H., RAVINDRANATH N., PAYNE S.,
RA   SEIDAH N.G.;
RT   "Molecular cloning and expression of the ovine testicular follicle
RT   stimulating hormone receptor.";
RL   MOL. CELL. ENDOCRINOL. 93:219-226(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 93176195.
RA   KHAN H., YARNEY T.A., SAIRAM M.R.;
```

```
RT   "Cloning of alternately spliced mRNA transcripts coding for variants
RT   of ovine testicular follitropin receptor lacking the G protein
RT   coupling domains.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 190:888-894(1993).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07302; G165885; -.
DR   PIR; JC1493; JC1493.
DR   GCRDB; GCR_0496; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367   387       1 (POTENTIAL).
FT   DOMAIN     388   398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399   421       2 (POTENTIAL).
FT   DOMAIN     422   443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444   465       3 (POTENTIAL).
FT   DOMAIN     466   485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486   508       4 (POTENTIAL).
FT   DOMAIN     509   528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529   550       5 (POTENTIAL).
FT   DOMAIN     551   573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574   597       6 (POTENTIAL).
FT   DOMAIN     598   608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609   630       7 (POTENTIAL).
FT   DOMAIN     631   695       CYTOPLASMIC (POTENTIAL).
FT   DISULFID   442   517       BY SIMILARITY.
FT   CARBOHYD   191   191       POTENTIAL.
FT   CARBOHYD   199   199       POTENTIAL.
FT   CARBOHYD   293   293       POTENTIAL.
SQ   SEQUENCE   695 AA;  78237 MW;  5C0B07D6 CRC32;

  Query Match          30.0%;  Score 219;  DB 1;  Length 695;
  Best Local Similarity  31.4%;  Pred. No. 1.77e-21;
  Matches   33;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQDSKVTEMPSDLPRDAVELRFVLTKLRVIPEGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |:|::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT  10
ID   FSHR_PIG       STANDARD;      PRT;   695 AA.
AC   P49059; O77514;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 96011644.
RA   REMY J.J., LAHBIB-MANSAIS Y., YERLE M., BOZON V., COUTURE L.,
RA   PAJOT E., GREBERT D., SALESSE R.;
RT   "The porcine follitropin receptor: cDNA cloning, functional
RT   expression and chromosomal localization of the gene.";
RL   GENE 163:257-261(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RA   WANG Y.F., MEYER K.B., SCHMIDT K., WAN S.J., DEGEN S.J.F.,
RA   LA BARBERA A.R.;
RT   "Porcine follicle-stimulating hormone receptor.";
RL   SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31966; G1066833; -.
DR   EMBL; AF025377; G3282511; -.
DR   GCRDB; GCR_1561; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367   387       1 (POTENTIAL).
FT   DOMAIN     388   398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399   421       2 (POTENTIAL).
FT   DOMAIN     422   443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444   465       3 (POTENTIAL).
FT   DOMAIN     466   485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486   508       4 (POTENTIAL).
FT   DOMAIN     509   528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529   550       5 (POTENTIAL).
FT   DOMAIN     551   573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574   597       6 (POTENTIAL).
FT   DOMAIN     598   608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609   630       7 (POTENTIAL).
FT   DOMAIN     631   695       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   191   191       POTENTIAL.
FT   CARBOHYD   199   199       POTENTIAL.
FT   CARBOHYD   293   293       POTENTIAL.
FT   DISULFID   442   517       BY SIMILARITY.
FT   CONFLICT     2     2       S -> A (IN REF. 1).
FT   CONFLICT    13    13       T -> S (IN REF. 1).
FT   CONFLICT    60    60       V -> A (IN REF. 1).
FT   CONFLICT   166   166       V -> M (IN REF. 1).
FT   CONFLICT   215   215       Q -> H (IN REF. 1).
FT   CONFLICT   247   247       K -> R (IN REF. 1).
FT   CONFLICT   257   257       S -> T (IN REF. 1).
FT   CONFLICT   334   334       D -> N (IN REF. 1).
FT   CONFLICT   349   349       E -> K (IN REF. 1).
```

```
FT   CONFLICT    352    352       T -> A (IN REF. 1).
FT   CONFLICT    383    383       V -> E (IN REF. 1).
FT   CONFLICT    407    407       A -> T (IN REF. 1).
FT   CONFLICT    421    421       V -> I (IN REF. 1).
FT   CONFLICT    427    427       T -> S (IN REF. 1).
FT   CONFLICT    435    435       D -> N (IN REF. 1).
FT   CONFLICT    483    483       L -> V (IN REF. 1).
FT   CONFLICT    550    550       T -> I (IN REF. 1).
FT   CONFLICT    586    586       A -> V (IN REF. 1).
FT   CONFLICT    607    607       S -> L (IN REF. 1).
FT   CONFLICT    691    691       R -> H (IN REF. 1).
SQ   SEQUENCE   695 AA;  78172 MW;  10A3EA81 CRC32;

  Query Match           30.0%;  Score 219;  DB 1;  Length 695;
  Best Local Similarity  32.4%;  Pred. No. 1.77e-21;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQESKVTEIPPDLPRNAVELRFVLTKLRVIPKGAFSGFGDLEKIEISQNDV 82
           | ||| :  |:   : |||::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
          : | :: |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   11
ID   FSHR_MACFA     STANDARD;      PRT;   695 AA.
AC   P32212;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 94071854.
RA   GROMOLL J., DANKBAR B., SHARMA R.S., NIESCHLAG E.;
RT   "Molecular cloning of the testicular follicle stimulating hormone
RT   receptor of the non human primate Macaca fascicularis and
RT   identification of multiple transcripts in the testis.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 196:1066-1072(1993).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X74454; G396802; -.
DR   PIR; S36452; S36452.
DR   PIR; JN0898; JN0898.
DR   GCRDB; GCR_0653; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN       18    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    367    387       1 (POTENTIAL).
FT   DOMAIN      388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    399    421       2 (POTENTIAL).
FT   DOMAIN      422    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    444    465       3 (POTENTIAL).
FT   DOMAIN      466    485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    486    508       4 (POTENTIAL).
FT   DOMAIN      509    528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    529    550       5 (POTENTIAL).
FT   DOMAIN      551    573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    574    597       6 (POTENTIAL).
FT   DOMAIN      598    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    630       7 (POTENTIAL).
FT   DOMAIN      631    695       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    191    191       POTENTIAL.
FT   CARBOHYD    199    199       POTENTIAL.
FT   CARBOHYD    293    293       POTENTIAL.
FT   CARBOHYD    318    318       POTENTIAL.
FT   DISULFID    442    517       BY SIMILARITY.
SQ   SEQUENCE   695 AA;  78343 MW;  ED3151A9 CRC32;

  Query Match           29.9%;  Score 218;  DB 1;  Length 695;
  Best Local Similarity  31.4%;  Pred. No. 2.72e-21;
  Matches   33;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNAIELRFVHTKLRVIQKGAFSGFGDLEKIEISQNDV 82
           | ||| :  |:   : |||::||   :|:|: :  :: |  |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPNLRYLLI 127
          : | :| |  |  |  : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   12
ID   FSHR_HUMAN     STANDARD;      PRT;   695 AA.
AC   P23945;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 91222171.
RA   MINEGISH T., NAKAMURA K., TAKAKURA Y., IBUKI Y., IGARASHI M.;
RT   "Cloning and sequencing of human FSH receptor cDNA.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 175:1125-1130(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 93246012.
RA   KELTON C.A., CHENG S.V., NUGENT N.P., SCHWEICKHARDT R.L.,
RA   ROSENTHAL J.L., OVERTON S.A., WANDS G.D., KUZEJA J.B., LUCHETTE C.A.,
RA   CHAPPEL S.C.;
RT   "The cloning of the human follicle stimulating hormone receptor and
RT   its expression in COS-7, CHO, and Y-1 cells.";
RL   MOL. CELL. ENDOCRINOL. 89:141-151(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   TILLY L.T., AIHARA T., NISHIMORI K., JAI X.-C., BILLIG H.,
RA   KOWALSKI K.I., PERLAS E.A., HSUEH A.J.;
RL   SUBMITTED (XXX-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE OF 1-342 FROM N.A.
RC   TISSUE=TESTIS;
```

```
RX   MEDLINE; 93075197.
RA   GROMOLL J., GUDERMANN T., NIESCHLAG E.;
RT   "Molecular cloning of a truncated isoform of the human follicle
RT   stimulating hormone receptor.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 188:1077-1083(1992).
RN   [5]
RP   SEQUENCE OF 1-51 FROM N.A.
RX   MEDLINE; 95011044.
RA   GROMOLL J., DANKBAR B., GUDERMANN T.;
RT   "Characterization of the 5' flanking region of the human follicle-
RT   stimulating hormone receptor gene.";
RL   MOL. CELL. ENDOCRINOL. 102:93-102(1994).
RN   [6]
RP   3D-STRUCTURE MODELLING OF 49-228.
RX   MEDLINE; 96363672.
RA   JIANG X., DREANO M., BUCKLER D.R., CHENG S., YTHIER A., WU H.,
RA   HENDRICKSON W.A., EL TAYAR N.;
RT   "Structural predictions for the ligand-binding region of glycoprotein
RT   hormone receptors and the nature of hormone-receptor interactions.";
RL   STRUCTURE 3:1341-1353(1995).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SERTOLI CELLS AND OVARIAN GRANULOSA CELLS.
CC   -!- ALTERNATIVE PRODUCTS: A SHORT FORM OF THE TESTICULAR PROTEIN IS
CC       PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE
CC       SHOWN HERE IS THAT OF THE LONG TESTICULAR PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M65085; G182771; -.
DR   EMBL; S59900; G300073; -.
DR   EMBL; M95489; G182773; -.
DR   EMBL; X68044; G31474; -.
DR   EMBL; S73199; G685037; -.
DR   PIR; JN0122; JN0122.
DR   PDB; 1XUN; 15-MAY-97.
DR   GCRDB; GCR_0071; -.
DR   GCRDB; GCR_0404; -.
DR   GCRDB; GCR_0588; -.
DR   GCRDB; GCR_0690; -.
DR   MIM; 136435; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT; ALTERNATIVE SPLICING;
KW   3D-STRUCTURE.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367    387       1 (POTENTIAL).
FT   DOMAIN     388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399    421       2 (POTENTIAL).
FT   DOMAIN     422    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444    465       3 (POTENTIAL).
FT   DOMAIN     466    485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486    508       4 (POTENTIAL).
FT   DOMAIN     509    528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529    550       5 (POTENTIAL).
FT   DOMAIN     551    573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574    597       6 (POTENTIAL).
FT   DOMAIN     598    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    630       7 (POTENTIAL).
FT   DOMAIN     631    695       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   191    191       POTENTIAL.
FT   CARBOHYD   199    199       POTENTIAL.
FT   CARBOHYD   293    293       POTENTIAL.
FT   CARBOHYD   318    318       POTENTIAL.
FT   DISULFID   442    517       BY SIMILARITY.
FT   VARSPLIC   224    285       MISSING (IN SHORT TESTICULAR FORM).
FT   VARSPLIC   342    695       MISSING (IN SHORT TESTICULAR FORM).
FT   CONFLICT    13     13       S -> R (IN REF. 4).
FT   CONFLICT   112    112       N -> T (IN REF. 1).
FT   CONFLICT   197    198       EL -> AV (IN REF. 1).
FT   CONFLICT   295    295       S -> P (IN REF. 4).
FT   CONFLICT   307    307       T -> A (IN REF. 1).
FT   CONFLICT   680    680       N -> S (IN REF. 1).
SQ   SEQUENCE   695 AA;  78294 MW;  7032FA16 CRC32;

  Query Match           29.7%;  Score 217;  DB 1;  Length 695;
  Best Local Similarity  32.4%;  Pred. No. 4.16e-21;
  Matches   34;  Conservative   26;  Mismatches  43;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQESKVTEIPSDLPRNAIELRFVLTKLRVIQKGAFSGFGDLEKIEISQNDV 82
           | ||| :  |:   : |||::||   :|:|:    :: |  |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPNLQYLLI 127
             : | :| |  |  |  : :  :|::  |   |: | :|| ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT  13
ID   FSHR_EQUAS     STANDARD;      PRT;   687 AA.
AC   Q95179;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE   RECEPTOR).
GN   FSHR.
OS   EQUUS ASINUS (DONKEY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PERISSODACTYLA; EQUIDAE; EQUUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 97338913.
RA   RICHARD F., MARTINAT N., REMY J.-J., SALESSE R., COMBARNOUS Y.;
RT   "Cloning, sequencing and in vitro functional expression of
RT   recombinant donkey follicle-stimulating hormone receptor: a new
RT   insight into the binding specificity of gonadotrophin receptors.";
RL   J. MOL. ENDOCRINOL. 18:193-202(1997).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGHLY SIMILAR TO LSH AND TSH RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U73659; G1658014; -.
DR   GCRDB; GCR_1116; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P23945; 1XUN.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW   PHOSPHORYLATION; REPEAT; LEUCINE-REPEAT.
```

```
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   687       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   358       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   359   379       1 (POTENTIAL).
FT   DOMAIN     380   390       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   391   413       2 (POTENTIAL).
FT   DOMAIN     414   435       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   436   457       3 (POTENTIAL).
FT   DOMAIN     458   477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   478   500       4 (POTENTIAL).
FT   DOMAIN     501   520       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   521   542       5 (POTENTIAL).
FT   DOMAIN     543   565       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   566   589       6 (POTENTIAL).
FT   DOMAIN     590   600       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   601   622       7 (POTENTIAL).
FT   DOMAIN     623   687       CYTOPLASMIC (POTENTIAL).
FT   DISULFID   434   509       BY SIMILARITY.
FT   CARBOHYD   191   191       POTENTIAL.
FT   CARBOHYD   199   199       POTENTIAL.
FT   CARBOHYD   293   293       POTENTIAL.
SQ   SEQUENCE   687 AA;  76937 MW;  53777553 CRC32;

  Query Match          29.5%;  Score 215;  DB 1;  Length 687;
  Best Local Similarity  32.4%;  Pred. No. 9.72e-21;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHYSNRVFLCQESKVTEIPSDLPRNALELRFVLTKLRVIPKGAFSGFGDLKKIEISQNDV 82
          |  ||  :  |:   : |||::||   :|:|:    :: || |||: :  ||:|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDHDAFQNLPNLQYLLI 127
             : | :: |  |  |  : :  :|::  |    |: | :|| ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLLL 105


RESULT   14
ID   PGS1_BOVIN     STANDARD;      PRT;   369 AA.
AC   P21809; P79259;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (LEUCINE-RICH PG I)
DE   (PG-S1).
GN   BGN.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AORTA;
RX   MEDLINE; 96113563.
RA   XU J.H., RADHAKRISHNAMURTHY B., SRINIVASAN S.R., BERENSON G.S.;
RT   "Primary structure of bovine aorta biglycan core protein deduced from
RT   cloned CDNA.";
RL   BIOCHEM. MOL. BIOL. INT. 37:263-272(1995).
RN   [2]
RP   SEQUENCE OF 38-369.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 89255324.
RA   NEAME P.J., CHOI H.U., ROSENBERG L.C.;
RT   "The primary structure of the core protein of the small, leucine-rich
RT   proteoglycan (PG I) from bovine articular cartilage.";
RL   J. BIOL. CHEM. 264:8653-8661(1989).
RN   [3]
RP   SEQUENCE OF 38-63.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 89123388.
RA   CHOI H.U., JOHNSON T.L., PAL S., TANG L.H., ROSENBERG L.C.,
RA   NEAME P.J.;
RT   "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT   DS-PGII, from bovine articular cartilage and skin isolated by octyl-
RT   sepharose chromatography.";
RL   J. BIOL. CHEM. 264:2876-2884(1989).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S82652; G1835865; -.
DR   PIR; A33701; A33701.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   SIGNAL; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1    19       POTENTIAL.
FT   PROPEP      20    37
FT   CHAIN       38   369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      93   316       LEUCINE-RICH REPEATS.
FT   REPEAT      93   106       LRR 1.
FT   REPEAT     117   130       LRR 2.
FT   REPEAT     141   154       LRR 3.
FT   REPEAT     162   175       LRR 4.
FT   REPEAT     186   199       LRR 5.
FT   REPEAT     211   224       LRR 6.
FT   REPEAT     232   245       LRR 7.
FT   REPEAT     256   269       LRR 8.
FT   REPEAT     280   288       LRR 9.
FT   REPEAT     303   316       LRR 10.
FT   CARBOHYD    42    42       GLYCOSAMINOGLYCAN.
FT   CARBOHYD    48    48       GLYCOSAMINOGLYCAN.
FT   CARBOHYD   271   271
FT   CARBOHYD   312   312
FT   DISULFID    64    77
FT   DISULFID   322   355
FT   CONFLICT   152   152       C -> V (IN REF. 2).
FT   CONFLICT   188   188       C -> E (IN REF. 2).
FT   CONFLICT   354   354       A -> R (IN REF. 2).
FT   CONFLICT   368   369       KK -> Y (IN REF. 2).
SQ   SEQUENCE   369 AA;  41509 MW;  F1CC673B CRC32;

  Query Match          28.9%;  Score 211;  DB 1;  Length 369;
  Best Local Similarity  38.4%;  Pred. No. 5.29e-20;
  Matches   33;  Conservative   24;  Mismatches  26;  Indels   3;  Gaps   2;

Db    153 EIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK-L 211
          |||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:||| |
Qy     19 EIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLKSL 76

Db    212 NYLRISEAKLTGIPKDLPETLNELHL 237
              | :   |:| |:| | : |  |:|
Qy     77 TSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   15
ID   PGS2_CHICK     STANDARD;      PRT;   357 AA.
AC   P28675;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
```

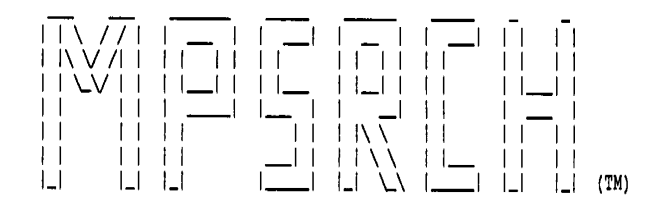
```
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=WHITE LEGHORN; TISSUE=CORNEA;
RX   MEDLINE; 92296755.
RA   LI W., VERGNES J.P., CORNUET P.K., HASSEL J.R.;
RT   "cDNA clone to chick corneal chondroitin/dermatan sulfate
RT   proteoglycan reveals identity to decorin.";
RL   ARCH. BIOCHEM. BIOPHYS. 296:190-197(1992).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X63797; G62888; -.
DR   PIR; S22197; S22197.
DR   PIR; S24317; S24317.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    16
FT   PROPEP      17    30
FT   CHAIN       31   357       BONE PROTEOGLYCAN II.
FT   DOMAIN      75   306       LEUCINE-RICH REPEATS.
FT   REPEAT      75    96       LRR 1.
FT   REPEAT      97   120       LRR 2.
FT   REPEAT     121   143       LRR 3.
FT   REPEAT     144   165       LRR 4.
FT   REPEAT     166   191       LRR 5.
FT   REPEAT     192   215       LRR 6.
FT   REPEAT     216   236       LRR 7.
FT   REPEAT     237   260       LRR 8.
FT   REPEAT     261   283       LRR 9.
FT   REPEAT     284   306       LRR 10.
FT   CARBOHYD    34    34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   209   209       POTENTIAL.
FT   CARBOHYD   260   260       POTENTIAL.
FT   DISULFID    52    65       BY SIMILARITY.
FT   DISULFID   311   344       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  39687 MW;  48F51E32 CRC32;

  Query Match          27.9%;  Score 204;  DB 1;  Length 357;
  Best Local Similarity  33.7%;  Pred. No. 1.01e-18;
  Matches   32;  Conservative   17;  Mismatches  46;  Indels   0;  Gaps   0;

Db     56 CQCHLRVVQCSDLGLERVPKDLPPDTTLLDLQNNKITEIKEGDFKNLKNLHALILVNNKI 115
          | |   :|:|   ||  :| :||     : |: | |  |  | |   |:|  : : :| |
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db    116 SKISPAAFAPLKKLERLYLSKNNLKELPENMPKSL 150
          | |:| ||  || |  | |  |:: |::  :  :|
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGL 97


Search completed: Fri May 28 08:43:33 1999
Job time : 22 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:43:51 1999; MasPar time 10.05 Seconds
570.514 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-3
Description: (1-105) from US09191647.pep
Perfect Score: 730
Sequence: 1 SPCTCSNNIVDCRGKGLMEI.....ITEIAKGLFDGLVSLQLLLL 105

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 41.576; Variance 84.632; scale 0.491

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	698	95.6	1523	11	O88280	MEGF5.	9.33e-113
2	578	79.2	1531	11	O88279	MEGF4.	9.68e-89
3	228	31.2	1496	4	Q92626	MYELOBLAST KIAA0230 (F	7.27e-22
4	221	30.3	369	6	O46390	BIGLYCAN PRECURSOR.	1.24e-20
5	221	30.3	688	11	Q64183	FOLLICLE-STIMULATING H	1.24e-20
6	220	30.1	372	6	O46403	BIGLYCAN.	1.86e-20
7	219	30.0	259	6	Q28574	FOLLITROPIN RECEPTOR P	2.79e-20
8	214	29.3	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	2.09e-19
9	210	28.8	134	6	Q28573	FOLLITROPIN RECEPTOR P	1.04e-18
10	210	28.8	516	4	O43300	KIAA0416.	1.04e-18
11	209	28.6	313	4	Q13288	P37NB.	1.55e-18
12	209	28.6	904	4	O15455	TOLL-LIKE RECEPTOR 3.	1.55e-18
13	191	26.2	360	6	Q28888	DECORIN.	1.90e-15
14	191	26.2	522	4	O43354	BAC CLONE GS099H08, CO	1.90e-15
15	190	26.0	184	4	O60803	DJ63G5.3 (PUTATIVE LEU	2.81e-15
16	190	26.0	360	6	O46542	DERMATAN SULFATE PROTE	2.81e-15
17	190	26.0	961	5	P90920	K07A12.2 PROTEIN.	2.81e-15
18	188	25.8	653	5	O02329	T23G11.6 PROTEIN.	6.13e-15
19	188	25.8	839	4	O00206	TOLL PROTEIN HOMOLOG.	6.13e-15
20	187	25.6	359	4	O15335	CHONDROADHERIN.	9.05e-15
21	185	25.3	135	6	O46377	BIGLYCAN (FRAGMENT).	1.97e-14
22	184	25.2	322	11	P70186	PROTEOGLYCAN PRECURSOR	2.90e-14
23	184	25.2	322	4	Q99645	DERMATAN SULFATE PROTE	2.90e-14
24	184	25.2	358	11	O55226	CHONDROADHERIN.	2.90e-14
25	183	25.1	321	6	P79119	EPIPHYCAN.	4.27e-14
26	183	25.1	603	11	O70211	INSULIN-LIKE GROWTH FA	4.27e-14
27	182	24.9	358	11	O70210	CHONDROADHERIN.	6.28e-14
28	180	24.7	316	13	Q90944	PROTEOGLYCAN CORE PROT	1.36e-13
29	180	24.7	680	5	Q93374	C44H4.3 PROTEIN.	1.36e-13
30	179	24.5	610	5	Q21604	M88.6 PROTEIN.	1.99e-13
31	177	24.2	428	4	O14498	ISLR PRECURSOR.	4.29e-13
32	177	24.2	1385	5	Q26388	TLR=TOLL-LIKE RECEPTOR	4.29e-13
33	175	24.0	353	13	O42235	KERATAN SULFATE PROTEO	9.23e-13
34	173	23.7	892	5	P91644	KEK2 PRECURSOR (FRAGME	1.98e-12
35	173	23.7	907	4	O75473	ORPHAN G PROTEIN-COUPL	1.98e-12
36	173	23.7	1389	5	Q24591	WHEELER.	1.98e-12
37	170	23.3	331	13	O93233	PHOSPHOLIPASE A2 INHIB	6.18e-12
38	168	23.0	716	11	Q61809	LEUCINE-RICH-REPEAT PR	1.32e-11
39	168	23.0	1066	5	Q18902	CODED FOR BY C. ELEGAN	1.32e-11
40	165	22.6	713	4	O75325	GLIOMA AMPLIFIED ON CH	4.07e-11
41	165	22.6	733	5	Q24250	TARTAN PROTEIN PRECURS	4.07e-11
42	163	22.3	352	6	O62702	CORNEAL KERATAN SULFAT	8.61e-11
43	163	22.3	352	6	Q28032	CORNEAL KERATAN SULFAT	8.61e-11
44	163	22.3	603	5	Q22075	T01G9.3 PROTEIN.	8.61e-11
45	160	21.9	784	4	O60603	TOLL/INTERLEUKIN-1 REC	2.64e-10

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 95.6%; Score 698; DB 11; Length 1523;
Best Local Similarity 94.3%; Pred. No. 9.33e-113;
Matches 99; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 282 SACSCSNNIVDCRGKGLTEIPANLPEGIVEIRLEQNSIKSIPAGAFIQYKKLKRIDISKN 341
|:|:|||||||||||||| ||||||||||||||||||||:|||||| ||||||||||||
Qy 1 SPCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKN 60

Db 342 QISDIAPDAFQGLKSLTSLVLYGNKITEIPKGLFDGLVSLQLLLL 386
|||||||||||||||||||||||||||||:|||||||||||||||
Qy 61 QISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105

RESULT 2
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match           79.2%;  Score 578;  DB 11;  Length 1531;
  Best Local Similarity  71.8%;  Pred. No. 9.68e-89;
  Matches   74;  Conservative   19;  Mismatches  10;  Indels   0;  Gaps   0;

Db    286 CSCSNGIVDCRGKGLTAIPANLPETMTEIRLELNGIKSIPPGAFSPYRKLRRIDLSNNQI 345
          |:||| |||||||||  ||||||| : ||||| |:||:||:|||: |:||:|||:|:|||
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db    346 AEIAPDAFQGLRSLNSLVLYGNKITDLPRGVFGGLYTLQLLLL 388
          ::|||||||||:|| ||||||||||::::|:| || :||||||
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    3
ID   Q92626      PRELIMINARY;      PRT;  1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELOBLAST KIAA0230 (FRAGMENT).
GN   KIAA0230.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA RES. 3:321-329(1996).
DR   EMBL; D86983; D1013908; -.
DR   PFAM; PF00047; ig; 4.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00141; peroxidase; 1.
DR   PFAM; PF00560; LRR; 3.
FT   NON_TER       1      1
SQ   SEQUENCE   1496 AA;  167209 MW;  5731EE51 CRC32;

  Query Match           31.2%;  Score 228;  DB 4;  Length 1496;
  Best Local Similarity  37.4%;  Pred. No. 7.27e-22;
  Matches   40;  Conservative   24;  Mismatches  38;  Indels   5;  Gaps   4;

Db     55 SRCLCFRTTVRCM-HLLLEAVPAVAPQTSILDLRF--NRIREIQPGAFRRLRNLNTLLLN 111
          | | |    | |    |:| :|| |: :|:::|:  | |: | :||| : ::|: : ::
Qy      1 SPCTCSNNIVDCRGKGLME-IPANLPE-GIVEIRLEQNSIKAIPAGAFTQYKKLKRIDIS 58

Db    112 NNQIKRIPSGAFEDLENLKYLYLYKNEIQSIDRQAFKGLASLEQLYL 158
          :||| |:: ||: | :|  | || | |  | :  | || ||: | |
Qy     59 KNQISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT    4
ID   O46390      PRELIMINARY;      PRT;  369 AA.
AC   O46390;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   BIGLYCAN PRECURSOR.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; CAPRINAE; OVIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CHOROID PLEXUS;
RA   BRUETT L., CLEMENTS J.E.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF034842; G2655356; -.
KW   SIGNAL.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       38    369       BIGLYCAN.
SQ   SEQUENCE   369 AA;  41523 MW;  A0A9F549 CRC32;

  Query Match           30.3%;  Score 221;  DB 6;  Length 369;
  Best Local Similarity  38.6%;  Pred. No. 1.24e-20;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    151 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLK 210
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    211 -LNYLRISEAKLTGIPKDLPETLNELHL 237
           | | :  |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT    5
ID   Q64183      PRELIMINARY;      PRT;  688 AA.
AC   Q64183;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FOLLICLE-STIMULATING HORMONE RECEPTOR.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92149579.
RA   HECKERT L.L., DALEY I.J., GRISWOLD M.D.;
RT   "Structural organization of the follicle-stimulating hormone receptor
RT   gene.";
RL   MOL. ENDOCRINOL. 6:70-80(1992).
DR   EMBL; S81119; E91753; JOINED.
DR   EMBL; S81121; E91753; JOINED.
DR   EMBL; S81171; E91753; JOINED.
DR   EMBL; S81174; E91753; JOINED.
DR   EMBL; S81198; E91753; -.
DR   EMBL; S81117; E91753; JOINED.
DR   EMBL; S81178; E91753; JOINED.
DR   EMBL; S81185; E91753; JOINED.
DR   EMBL; S81194; E91753; JOINED.
DR   EMBL; S81183; E91753; JOINED.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00560; LRR; 3.
SQ   SEQUENCE   688 AA;  77341 MW;  E454B260 CRC32;

  Query Match           30.3%;  Score 221;  DB 11;  Length 688;
  Best Local Similarity  32.4%;  Pred. No. 1.24e-20;
  Matches   34;  Conservative   27;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNRVFLCQDSKVTEIPTDLPRNAIELRFVLTKLRVIPKGSFAGFGDLEKIEISQNDV 82
```

```
          | ||| :  |:   : |||::||   :|:|:    :: || |:|: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLPSLRYLLI 127
           : | :| |  |  |  : :  :|:: |    |: | ||: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   6
ID   O46403      PRELIMINARY;      PRT;   372 AA.
AC   O46403;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   BIGLYCAN.
OS   EQUUS CABALLUS (HORSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PERISSODACTYLA; EQUIDAE; EQUUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RICHARDSON D.W., DODGE G.R.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF035934; G2662531; -.
SQ   SEQUENCE   372 AA;  41924 MW;  097E7BA6 CRC32;

  Query Match          30.1%;  Score 220;  DB 6;  Length 372;
  Best Local Similarity  38.6%;  Pred. No. 1.86e-20;
  Matches   34;  Conservative   25;  Mismatches  26;  Indels   3;  Gaps   2;

Db    154 LVEIPPNLPSSLVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFQPGAFDGLK 213
          |:|||:||| ::||:|: :| |: :| | |:  :::: |::: | :  | : | ||:|||
Qy     17 LMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI--SDIAPDAFQGLK 74

Db    214 -LNYLRISEAKLTGIPKDLPETLNELHL 240
           | | :   |:| |:| | : |  |:|
Qy     75 SLTSLVLYGNKITEIAKGLFDGLVSLQL 102


RESULT   7
ID   Q28574      PRELIMINARY;      PRT;   259 AA.
AC   Q28574;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FOLLITROPIN RECEPTOR PRECURSOR.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; CAPRINAE; OVIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DORSET-LEICESTER-SUFFOLK; TISSUE=TESTIS;
RX   MEDLINE; 93176195.
RA   KHAN H., YARNEY T.A., SAIRAM M.R.;
RT   "Cloning of alternately spliced mRNA transcripts coding for variants
RT   of ovine testicular follitropin receptor lacking the G protein
RT   coupling domains.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 190:888-894(1993).
DR   EMBL; L12767; G484229; -.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18     59       FOLLITROPIN RECEPTOR.
SQ   SEQUENCE   259 AA;  29095 MW;  7095A385 CRC32;

  Query Match          30.0%;  Score 219;  DB 6;  Length 259;
  Best Local Similarity  31.4%;  Pred. No. 2.79e-20;
  Matches   33;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;

Db     23 CHCSNGVFLCQDSKVTEMPSDLPRDAVELRFVLTKLRVIPEGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |:|::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLLI 127
           : | :: |  |  | : :  :|::  |    |: | :|: ||:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 105


RESULT   8
ID   P70193      PRELIMINARY;      PRT;  1091 AA.
AC   P70193;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96394313.
RA   SUZUKI Y., SATO N., TOHYAMA M., WANAKA A., TAKAGI T.;
RT   "cDNA cloning of a novel membrane glycoprotein that is expressed
RT   specifically in glial cells in the mouse brain LIG-1: A protein with
RT   leucine-rich repeats and immunoglobulin-like domains.";
RL   J. BIOL. CHEM. 271:22522-22527(1996).
DR   EMBL; D78572; D1012081; -.
DR   MGD; MGI:107935; IMG.
DR   PFAM; PF00047; ig; 3.
DR   PFAM; PF00560; LRR; 7.
KW   MEMBRANE.
SQ   SEQUENCE   1091 AA;  119283 MW;  C0F262F9 CRC32;

  Query Match          29.3%;  Score 214;  DB 11;  Length 1091;
  Best Local Similarity  42.5%;  Pred. No. 2.09e-19;
  Matches   37;  Conservative   19;  Mismatches  27;  Indels   4;  Gaps   3;

Db    330 AELS-SLSILRLSHNAISHIAEGAFKGLKSLRVLDLDHNEISGTIEDTSGAFTGLDNLSK 388
          |:|: ::  :|| :|:|  |: |||   | |: :|:  |:|| |  :: || || :|:
Qy     22 ANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQISD-I--APDAFQGLKSLTS 78

Db    389 LTLFGNKIKSVAKRAFSGLESLEHLNL 415
           | |:||||  :||  | || ||: | |
Qy     79 LVLYGNKITEIAKGLFDGLVSLQLLL 105


RESULT   9
ID   Q28573      PRELIMINARY;      PRT;   134 AA.
AC   Q28573;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FOLLITROPIN RECEPTOR PRECURSOR.
GN   SHPFSHRE.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; CAPRINAE; OVIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DORSET-LEICESTER-SUFFOLK; TISSUE=TESTIS;
RX   MEDLINE; 93176195.
RA   KHAN H., YARNEY T.A., SAIRAM M.R.;
RT   "Cloning of alternately spliced mRNA transcripts coding for variants
RT   of ovine testicular follitropin receptor lacking the G protein
RT   coupling domains.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 190:888-894(1993).
DR   EMBL; L12766; G484227; -.
DR   PFAM; PF00560; LRR; 1.
KW   SIGNAL.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    134       POTENTIAL.
SQ   SEQUENCE   134 AA;  15336 MW;  4D2F9627 CRC32;

  Query Match          28.8%;  Score 210;  DB 6;  Length 134;
  Best Local Similarity  30.8%;  Pred. No. 1.04e-18;
  Matches   32;  Conservative   28;  Mismatches  42;  Indels   2;  Gaps   2;
```

```
Db     23 CHCSNGVFLCQDSKVTEMPSDLPRDAVELRFVLTKLRVIPEGAFSGFGDLEKIEISQNDV 82
          | ||| :  |:   : |:|::||   ||:|:    :: || |||: :  | :|:|| |::
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     83 LEVIEANVFSNLPKLHEIRIEKANNLLYIDPDAFQNLPNLRYLF 126
          : | :: |  |  |  : :  :|::  |    |: | :|: |:
Qy     63 SD-IAPDAFQGLKSLTSLVLY-GNKITEIAKGLFDGLVSLQLLL 104


RESULT  10
ID   O43300      PRELIMINARY;      PRT;   516 AA.
AC   O43300;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   KIAA0416.
GN   KIAA0416.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   ISHIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB007876; D1025768; -.
SQ   SEQUENCE   516 AA;  59076 MW;  8EDFDB28 CRC32;

  Query Match           28.8%;  Score 210;  DB 4;  Length 516;
  Best Local Similarity  32.0%;  Pred. No. 1.04e-18;
  Matches   33;  Conservative   22;  Mismatches  48;  Indels   0;  Gaps   0;

Db     38 CRCEKLLFYCDSQGFHSVPNATDKGSLGLSLRHNHITELERDQFASFSQLTWLHLDHNQI 97
          | | : :  | :|:  :|     | : : | :| |  :    |: :  | :: :  |||
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     98 STVKEDAFQGLYKLKELILSSNKIFYLPNTTFTQLINLQNLDL 140
          | :  ||||||  |  |:| :|||  :::  |  |::|| | |
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105


RESULT  11
ID   Q13288      PRELIMINARY;      PRT;   313 AA.
AC   Q13288;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   P37NB.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97136875.
RA   KIM D., LAQUAGLIA M.P., YANG S.Y.;
RT   "A cDNA encoding a putative 37 kDa leucine-rich repeat (LRR) protein,
RT   p37NB, isolated from S-type neuroblastoma cell has a differential
RT   tissue distribution.";
RL   BIOCHIM. BIOPHYS. ACTA 1309:183-188(1996).
DR   EMBL; U32907; G1236329; -.
DR   PFAM; PF00560; LRR; 3.
SQ   SEQUENCE   313 AA;  36287 MW;  AD538A24 CRC32;

  Query Match           28.6%;  Score 209;  DB 4;  Length 313;
  Best Local Similarity  32.9%;  Pred. No. 1.55e-18;
  Matches   28;  Conservative   27;  Mismatches  30;  Indels   0;  Gaps   0;

Db     65 LDCQERKLVYVLPGWPQDLLHMLLARNKIRTLKNNMFSKFKKLKSLDLQQNEISKIESEA 124
          :||: : |: : :  |: :: : | :| |:::    |: :|||| :|:  |:|| | ::|
Qy     10 VDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQISDIAPDA 69

Db    125 FFGLNKLTTLLLQHNQIKVLTEEVF 149
          | ||: ||:|:|  | |  ::  :|
Qy     70 FQGLKSLTSLVLYGNKITEIAKGLF 94


RESULT  12
ID   O15455      PRELIMINARY;      PRT;   904 AA.
AC   O15455;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TOLL-LIKE RECEPTOR 3.
GN   TLR3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROCK F.L., HARDIMAN G., TIMANS J.C., KASTELEIN R.A., BAZAN J.F.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88879; G2459626; -.
DR   PFAM; PF00560; LRR; 12.
SQ   SEQUENCE   904 AA;  103828 MW;  F857CE1C CRC32;

  Query Match           28.6%;  Score 209;  DB 4;  Length 904;
  Best Local Similarity  32.7%;  Pred. No. 1.55e-18;
  Matches   34;  Conservative   26;  Mismatches  41;  Indels   3;  Gaps   3;

Db     28 CTVSHEVADCSHLKLTQVPDDLPTNITVLNLTHNQLRRLPAANFTRYSQLTSLDVGFNTI 87
          || |::: ||   .| ::| :||  |  : | :| :: :||: ||:|  |  :|:: | |
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db     88 SKLEPELCQKLPMLKVLNLQHNELSQLSDKTFAFCTNLTELHLM 131
          | : |:  | |  |  | |  | ::::: | : |   |  |:|:
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIA-KGL-F-DGLVSLQLL 103


RESULT  13
ID   Q28888      PRELIMINARY;      PRT;   360 AA.
AC   Q28888; Q28608;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DECORIN.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95122319.
RA   ZHAN Q., BURROWS R., CINTRON C.;
RT   "Cloning and in situ hybridization of rabbit decorin in corneal
RT   tissues.";
RL   INVEST. OPHTHALMOL. VIS. SCI. 36:206-215(1995).
RN   [2]
RP   SEQUENCE OF 38-358 FROM N.A.
RC   TISSUE=CARTILAGE;
RA   HERING T.M., KOLLAR J.;
RL   SUBMITTED (NOV-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; S76584; G913375; -.
DR   EMBL; U03394; G415647; -.
DR   PFAM; PF00560; LRR; 5.
SQ   SEQUENCE   360 AA;  39896 MW;  1FAB2F8E CRC32;

  Query Match           26.2%;  Score 191;  DB 6;  Length 360;
  Best Local Similarity  32.3%;  Pred. No. 1.90e-15;
  Matches   32;  Conservative   17;  Mismatches  50;  Indels   0;  Gaps   0;

Db     59 CQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHALILVNNKI 118
          | |  :|:|  || :| :||    : |: | |  |  | |  |:|  : : :| |
Qy      3 CTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQI 62

Db    119 SKISPGAFTPLVKLERLYLSKNHLKELPEKMPKSLQELR 157
          | |:| ||  |  |  | |  | : |::  :  :|  |:
```

```
Qy     63 SDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQ 101


RESULT  14
ID   O43354      PRELIMINARY;      PRT;   522 AA.
AC   O43354;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   BAC CLONE GS099H08, COMPLETE SEQUENCE.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROHLFING T., DAVID M., AHRENS C.;
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
[redacted]
RN   [2]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AC004010; G2781386; -.
SQ   SEQUENCE   522 AA;  57933 MW;  0E8FD1F7 CRC32;

  Query Match          26.2%;  Score 191;  DB 4;  Length 522;
  Best Local Similarity  29.2%;  Pred. No. 1.90e-15;
  Matches   31;  Conservative   27;  Mismatches  47;  Indels   1;  Gaps   1;

Db     43 TACICATDIVSCTNKNLSKVPGNLFRLIKRLDLSYNRIGLLDSEWIPVSFAKLNTLILRH 102
           ::| |: :|| |  | |  :|:||   |  : |  | |  : :  :   : ||: : :
Qy      1 SPCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFT-QYKKLKRIDISK 59

Db    103 NNITSISTGSFSTTPNLKCLDLSSNKLKTVKNAVFQELKVLEVLLL 148
          | |: |:  :|    :|  | | :||:  : :::|: |  |::|||
Qy     60 NQISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLL 105

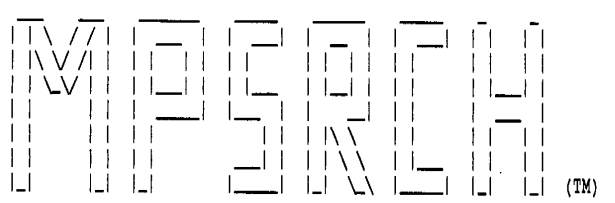

RESULT  15
ID   O60803      PRELIMINARY;      PRT;   184 AA.
AC   O60803;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   DJ63G5.3 (PUTATIVE LEUCINE RICH PROTEIN) (FRAGMENT).
GN   DJ63G5.3.
[redacted]
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LLOYD D.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z94160; E1296586; -.
FT   NON_TER       1      1
FT   NON_TER     184    184
SQ   SEQUENCE   184 AA;  20630 MW;  7F359903 CRC32;

  Query Match          26.0%;  Score 190;  DB 4;  Length 184;
  Best Local Similarity  32.2%;  Pred. No. 2.81e-15;
  Matches   28;  Conservative   31;  Mismatches  26;  Indels   2;  Gaps   2;

Db      6 IPQHINSTVHDLRLNENKLKAVLYSSLNRFGNLTDLNLTKNEISYIEDGAFLGQSSLQVL 65
            || ::   : ::||::| :||:  ::: :: :|  ::::||:|| |   || |  ||  |
Qy     20 IPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLKRIDISKNQISDIAPDAFQGLKSLTSL 79

Db     66 QLGY-NKLSNLTEGMLRGMSRLQFLFV 91
           | | ||::::: |:: |:  ||:|::
Qy     80 VL-YGNKITEIAKGLFDGLVSLQLLLL 105


Search completed: Fri May 28 08:44:23 1999
Job time : 32 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:46:09 1999; MasPar time 7.71 Seconds
380.559 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-4
Description: (1-138) from US09191647.pep
Perfect Score: 995
Sequence: 1 EGAFNGAASVQELMLTGNQL.....CQKPFFLKEIPIQGVGHPGI 138

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 30.450; Variance 125.613; scale 0.242

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	678	68.1	1534	30	W46966	Amino acid sequence o	6.89e-56
2	322	32.4	1480	5	R25079	Drosophila SLIT prote	1.88e-20
3	246	24.7	196	5	R29102	Drosophila SLIT prote	3.47e-13
4	217	21.8	560	12	R71294	Human glycoprotein V.	1.79e-10
5	208	20.9	1091	27	W41641	Sequence used in dete	1.22e-09
6	184	18.5	605	17	R85888	WD-40 domain-contg. i	1.90e-07
7	179	18.0	603	17	R85889	WD-40 domain-contg. r	5.39e-07
8	175	17.6	245	28	W39170	Human PKD1 protein fr	1.23e-06
9	175	17.6	455	28	W39171	Human PKD1 protein fr	1.23e-06
10	175	17.6	4302	29	W33396	Human PKD1 polypeptid	1.23e-06
11	175	17.6	4302	28	W23830	Human PKD1 protein.	1.23e-06
12	175	17.6	4302	19	W00870	Polycystic kidney dis	1.23e-06
13	175	17.6	4339	15	R75916	Polycystic kidney dis	1.23e-06
14	175	17.6	4339	19	R87539	Polycystic kidney dis	1.23e-06
15	167	16.8	4303	17	R90302	Polycystic kidney dis	6.42e-06
16	165	16.6	345	23	W09405	Pineal gland specific	9.67e-06
17	155	15.6	784	39	W86350	Human DNAX toll-like	7.42e-05
18	155	15.6	784	39	W90069	Human TNF-alpha conve	7.42e-05
19	155	15.6	784	29	W48245	Human pro-tumour necr	7.42e-05
20	154	15.5	661	28	W47274	Human B-cell activati	9.08e-05
21	151	15.2	661	39	W87556	B cell surface protei	1.66e-04
22	151	15.2	661	25	W28510	Product of clone J422	1.66e-04
23	140	14.1	1025	18	W03185	Rice Xa21 disease res	1.50e-03
24	138	13.9	293	1	P91368	45 kDa amino terminal	2.23e-03
25	137	13.8	365	39	W86353	Partial human DNAX to	2.72e-03
26	136	13.7	1112	16	R85299	Tomato pathogen resis	3.31e-03
27	136	13.7	1112	16	R85298	Tomato pathogen resis	3.31e-03
28	135	13.6	139	8	R42263	Decorin sequence PT-7	4.03e-03
29	135	13.6	186	8	R42264	Decorin sequence PT-7	4.03e-03
30	135	13.6	234	8	R42265	Decorin sequence PT-7	4.03e-03
31	135	13.6	280	8	R42266	Decorin sequence PT-7	4.03e-03
32	135	13.6	305	8	R42267	Decorin sequence PT-7	4.03e-03
33	135	13.6	331	8	R42260	Mature decorin PT-65.	4.03e-03
34	135	13.6	342	17	R89439	Human recombinant dec	4.03e-03
35	135	13.6	353	1	R05160	Sequence of human bon	4.03e-03
36	135	13.6	1388	18	R89471	Collagen/decorin fusi	4.03e-03
37	132	13.3	904	39	W86351	Human DNAX toll-like	7.26e-03
38	128	12.9	644	39	W82318	Human 7-transmembrane	1.58e-02
39	128	12.9	799	39	W86352	Human DNAX toll-like	1.58e-02
40	128	12.9	806	20	W09254	Tomato pathogen resis	1.58e-02
41	128	12.9	806	16	R85301	Tomato pathogen resis	1.58e-02
42	128	12.9	837	39	W86361	Human DNAX toll-like	1.58e-02
43	128	12.9	863	14	R75919	Tomato Cf-9.	1.58e-02
44	127	12.8	332	15	R87953	Bovine neurotrophic b	1.92e-02
45	125	12.6	610	23	W18201	Platelet glycoprotein	2.83e-02

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 68.1%; Score 678; DB 30; Length 1534;
Best Local Similarity 66.7%; Pred. No. 6.89e-56;
Matches 92; Conservative 28; Mismatches 16; Indels 2; Gaps 2;

Db 582 dgafegaasvselhltanqlesirsgmfrg-ldglrtlmlrnnriscihndsftglrnvr 640
:|||:||||| || ||:||||:::: ||| | ||:|||||:| |:|: ||:|:|| :||
Qy 1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db 641 llslydnqittvspgafdtlqslstlnllanpfncncql-awlggwlrkrkivtgnprcq 699
|||||||:|||::|||| || ||||:|||:|||||||:| | || |||||:||:||||||

```
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQ 120

Db    700 npdflrqiplqdvafpdf 717
          :| ||::||:| |: | :
Qy    121 KPFFLKEIPIQGVGHPGI 138


RESULT    2
ID    R25079 standard; Protein; 1480 AA.
AC    R25079;
DT    05-JAN-1993  (first entry)
DE    Drosophila SLIT protein involved in axon pathway development.
KW    Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW    embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW    midline glial cells; axonogenesis; cell-cell interaction; ss.
OS    Drosophila melanogaster.
FH    Key             Location/Qualifiers
FT    peptide         1..36
FT                    /label= signal
FT    domain          73..294
FT                    /label= Flank_LRR_Flank_1
FT                    /note= "mediates adhesive events"
FT    domain          295..518
FT                    /label= Flank-LRR-Flank_2
FT                    /note= "mediates adhesive events"
FT    domain          519..714
FT                    /label= Flank_LRR_Flank_3
FT                    /note= "mediates adhesive events"
FT    domain          715..910
FT                    /label= Flank_LRR_Flank_4
FT                    /note= "mediates adhesive events"
FT    region          911..1150
FT                    /label= Tandem_EGF_like_repeats
FT                    /note= "involved in protein-protein interactions"
FT    region          1353..1393
FT                    /label= 7th_EGF_like_repeat
FT                    /note= "involved in receptor-ligand interactions"
FT    region          1394..1404
FT                    /label= alternative_splice_segment
FT                    /note= "developmentally regulated"
FT    region          1405..1480
FT                    /label= C-terminal_region
PN    WO9210518-A.
PD    25-JUN-1992.
PF    27-NOV-1991; U09055.
PR    07-DEC-1990; US-624135.
PA    (UYYA ) UNIV YALE.
PI    Artavanis-Tsakonas S, Rothberg JM;
DR    WPI; 92-234590/28.
DR    N-PSDB; Q25811.
PT    SLIT protein and sequence elements for treating
PT    neuro-degenerative disease - useful for Alzheimer's disease,
PT    nerve damage and Parkinson's disease, for diagnosis of cancer
PS    Claim 1; Page 84-89; 122pp; English.
CC    The SLIT protein is necessary for normal development of the midline
CC    of the CNS, partic. the midline glial cells, and for the
CC    concomitant formation of the commisural axon pathways. The process
CC    is dependent on the level of SLIT protein expression. It appears
CC    that SLIT protein is excreted by the midline glial cells where it
CC    is synthesised and is eventually associated with the surfaces of
CC    axons that traverse tham. The SLIT protein is tightly localised to
CC    the muscle attachment sites and to the sites of contact between
CC    adjacent pairs of cardioblasts as they coalesce to form the lumen
CC    of the larval heart. The SLIT protein defines a new set of
CC    molecules (TAGONS) which play a key role in axon outgrowth and
CC    pathfinding. SLIT can be used as a nerve regenerative in
CC    neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC    injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC    sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC    strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC    degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC    region-Flank" domains, the C-terminal region and part of the
CC    alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC    claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC    and at least 1 EGF-like repeat element from the SLIT protein.
CC    See also R29102.
SQ    Sequence   1480 AA;

  Query Match          32.4%;  Score 322;  DB 5;  Length 1480;
  Best Local Similarity  36.7%;  Pred. No. 1.88e-20;
  Matches   44;  Conservative   36;  Mismatches  38;  Indels   2;  Gaps   2;

Db    365 alsglkqlttlvlygnkikdlpsgvfkg-lgslrllllnaneiscirkdafrdlhslsll 423
          |::|   :  |:| || :  : :  |:| |::|: |:| :| |:|: :|:|  | |: ||
Qy      3 AFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLL 62

Db    424 slydnniqsvangtfdamksmktvhlaknpficdcnlrw-vadylhknpietsgarcesp 482
          ||||| | ::: |:| :: |: |::|  ||| |:|:|   ::  |:|  | :: :||: |
Qy     63 SLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT    3
ID    R29102 standard; Protein; 196 AA.
AC    R29102;
DT    05-JAN-1993  (first entry)
DE    Drosophila SLIT protein Flank-LRR-Flank consensus sequence.
KW    Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW    embryonic CNS; leucine-rich repeat; midline glial cells;
KW    axonogenesis; cell-cell interaction; ss.
OS    Drosophila melanogaster.
FH    Key             Location/Qualifiers
FT    region          1..32
FT                    /label= amino_flanking_region
FT    region          33..135
FT                    /label= Leucine_rich_repeat_region
FT    region          136..196
FT                    /label= carboxy_flanking_region
PN    WO9210518-A.
PD    25-JUN-1992.
PF    27-NOV-1991; U09055.
PR    07-DEC-1990; US-624135.
PA    (UYYA ) UNIV YALE.
PI    Artavanis-Tsakonas S, Rothberg JM;
DR    WPI; 92-234590/28.
PT    SLIT protein and sequence elements for treating
PT    neuro-degenerative disease - useful for Alzheimer's disease,
PT    nerve damage and Parkinson's disease, for diagnosis of cancer
PS    Claim 5; Page 95; 122pp; English.
CC    The SLIT protein is necessary for normal development of the midline
CC    of the CNS, partic. the midline glial cells, and for the
CC    concomitant formation of the commisural axon pathways. The process
CC    is dependent on the level of SLIT protein expression. It appears
CC    that SLIT protein is excreted by the midline glial cells where it
CC    is synthesised and is eventually associated with the surfaces of
CC    axons that traverse tham. The SLIT protein is tightly localised to
CC    the muscle attachment sites and to the sites of contact between
CC    adjacent pairs of cardioblasts as they coalesce to form the lumen
CC    of the larval heart. The SLIT protein defines a new set of
CC    molecules (TAGONS) which play a key role in axon outgrowth and
CC    pathfinding. SLIT contains 4 Flank-LRR-Flank domains (see R25079)
CC    which mediate adhesive events and this consensus sequence is based
CC    on them. Each of the 4 individual domain sequences is individually
CC    claimed. See also Q25811.
SQ    Sequence   196 AA;

  Query Match          24.7%;  Score 246;  DB 5;  Length 196;
  Best Local Similarity  23.0%;  Pred. No. 3.47e-13;
  Matches   31;  Conservative   12;  Mismatches  90;  Indels   2;  Gaps   2;

Db     48 fxxlxxlxxlxlxxnxixxlxxxxfxx-lxxlxxlilxxnxixxlxxxxlxxlxxlxxlx 106
          |    : | |  | :  :    |   | |  |:|  | |  :    :  |  :  |
Qy      4 FNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLS 63

Db    107 lxxnxixxlxxxxfxxlxxlxxlxxnpfxcdcxlxw-lxxxxxxxxxxxxxxrcxxpx 165
          |  | |  :    |  |  |  : |  ||| |:| |   |              ||  |
Qy     64 LYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKPF 123
```

```
Db    166 xxxxxxixxlxxxxf 180
              |  :   :
Qy    124 FLKEIPIQGVGHPGI 138


RESULT    4
ID   R71294 standard; Protein; 560 AA.
AC   R71294;
DT   18-AUG-1995  (first entry)
DE   Human glycoprotein V.
KW   Glycoprotein V; GPV; platelet.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..16
FT                   /label= Sig_peptide
FT   modified_site   51
FT                   /label= N-glycosylation_site
FT   modified_site   181
FT                   /label= N-glycosylation_site
FT   modified_site   244
FT                   /label= N-glycosylation_site
FT   modified_site   267
FT                   /note= "N-glycosylation_site"
FT   modified_site   298
FT                   /label= N-glycosylation_site
FT   modified_site   312
FT                   /label= N-glycosylation_site
FT   modified_site   385
FT                   /label= N-glycosylation_site
FT   cleavage_site   476..477
FT                   /note= "putative thrombin cleavage site"
FT   modified_site   499
FT                   /label= N-glycosylation_site
FT   domain          520..544
FT                   /note= "putative transmembrane domain"
PN   WO9502054-A.
PD   19-JAN-1995.
PF   07-JUL-1994; U07644.
PR   09-JUL-1993; US-089455.
PR   03-DEC-1993; US-162599.
PR   10-FEB-1994; US-195006.
PA   (CORT-) COR THERAPEUTICS INC.
PI   Cazenave J,  Lanza F,  Phillips DR;
DR   WPI; 95-066899/09.
DR   N-PSDB; Q85594.
PT   Platelet glycoprotein V gene - useful for producing glycoprotein
PT   V (GPV) and variants and generating antibodies to GPV
PS   Disclosure; Page 45-50; 82pp; English.
CC   Genomic clones were isolated from a human fibroblast library in
CC   lambda Fix using a 748 bp 32P-labeled glycoprotein V (GPV) cDNA
CC   probe.  Exon-containing fragments from positive clones were
CC   subcloned and sequenced.  The full sequence of the human GPV gene
CC   is given in Q85594.
SQ   Sequence   560 AA;

  Query Match          21.8%;  Score 217;  DB 12;  Length 560;
  Best Local Similarity  34.4%;  Pred. No. 1.79e-10;
  Matches   43;  Conservative   28;  Mismatches  49;  Indels   5;  Gaps   4;

Db    332 qgafqqlgelqvlalhsnqltalpdqllrg-lgklrqvslrrnrlralpralfrnlssle 390
          :||| | : :| | | :| | ::    :|| |: |: : || | :  ::   |  |||:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    391 svqldhnqletlpgdvfgalprltevllghnswrcdcglgpflg-wlrqhlglvggeepp 449
          : |  |:: |:    | :|  |: : |  |:  |:| ||: || ||| :  :|:|:  |
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRR-IVSGN--P 117

Db    450 rcagp 454
            ||  |
Qy    118 RCQKP 122


RESULT    5
ID   W41641 standard; Protein; 1091 AA.
AC   W41641;
DT   27-APR-1998  (first entry)
DE   Sequence used in detection method.
KW   Detection; mouse; murine.
OS   Mus sp.
PN   J09107971-A.
PD   28-APR-1997.
PF   19-OCT-1995; 270822.
PR   19-OCT-1995; JP-270822.
PA   (TANA ) TANABE SEIYAKU CO.
PA   (TOYA/) TOYAMA S.
DR   WPI; 97-292464/27.
DR   N-PSDB; V04445.
PT   Detection of genes, useful for cloning genes of high and low
PT   expression - by homogenising prepared ds-cDNA pool from sample for
PT   comparison with each other to remove specific DNA fragment
PS   Claim 9; Pages 13-16; 18pp; Japanese.
CC   The present sequence was used in the development of a novel method
CC   for detecting genes showing difference in expression quantity
CC   between several samples. The method comprises preparing a double
CC   stranded cDNA pool of a standard organism sample, and homogenising
CC   the contents of each DNA fragment contained in the resultant cDNA
CC   pool to prepare a content homogenised standard cDNA pool. Double
CC   stranded cDNA pools derived from each sample for several organism
CC   samples are compared with each other, and the DNA fragment
CC   associated with the DNA fragment in the cDNA pool from the content
CC   homogenised standard cDNA pool is removed to give a remaining cDNA
CC   pool for each of the samples to be compared. The method can clone
CC   gene groups of high and low expression level efficiently.
SQ   Sequence   1091 AA;

  Query Match          20.9%;  Score 208;  DB 27;  Length 1091;
  Best Local Similarity  33.1%;  Pred. No. 1.22e-09;
  Matches   45;  Conservative   30;  Mismatches  57;  Indels   4;  Gaps   4;

Db    350 egafkglkslrvldldhneisgtiedtsgaftgldnlskltlfgnkiksvakrafsgles 409
          ||||:|  |:: | |  |:: |:    |  ||  |  | | :| |  |:: :|:|| |
Qy      1 EGAFNGAASVQELMLTGNQLE-TVHG-RGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSS 58

Db    410 lehlnlgenairsvqfdafakmknlkelyissesflcdcqlkw-lppwlmgrmlqafvta 468
          :  |:| :| | ::   ||: : :|  : : |::| |:|:|   |  ||  | : :   :
Qy     59 VRLLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVS-GNP 117

Db    469 tcahpeslkgqsifsv 484
            |  |  ||  :| :|
Qy    118 RCQKPFFLKEIPIQGV 133


RESULT    6
ID   R85888 standard; Protein; 605 AA.
AC   R85888;
DT   13-SEP-1996  (first entry)
DE   WD-40 domain-contg. insulin-like growth factor binding protein.
KW   WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW   intracellular signalling; protein kinase C; homology; motif; modulator;
KW   receptors of activated protein kinase; enzyme activity; isozyme; human.
OS   Synthetic.
PN   WO9521252-A2.
PD   10-AUG-1995.
PF   31-JAN-1995; U01210.
PR   01-FEB-1994; US-190802.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Mochly-Rosen D,  Ron D;
DR   WPI; 95-283772/37.
PT   New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT   activity of a protein, eg. protein kinase C, which interacts with a
PT   protein contg. a WD-40 region.
PS   Example 5; Page 122-125; 351pp; English.
CC   Proteins R85851-92 are protein which contain at least one WD-40 (also
CC   called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC   regions are involved in protein-protein interactions between proteins
```

```
CC   involved in intracellular signalling. An example of such an interaction
CC   is between protein kinase C and receptors of activated protein kinase
CC   (RACK), esp. RACK-1 (R85850). Proteins R85851-82 were isolated based on
CC   homology with beta-transducin, whereas proteins R85882-92 were isolated
CC   based on homology with the WD-40 consensus sequence (R85893). The
CC   proteins were used to construct the peptides R84928-R85063 and
CC   R85786-R85842. The peptides can be used to identify target proteins
CC   contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC   proteins involved in protein-protein interaction and to screen for drugs
CC   that will affect protein-protein interaction involving WD-40 domains.
SQ   Sequence  605 AA;

  Query Match          18.5%;  Score 184;  DB 17;  Length 605;
  Best Local Similarity  34.8%;  Pred. No. 1.90e-07;
  Matches   31;  Conservative   27;  Mismatches  28;  Indels   3;  Gaps   3;

Db    187 daafrglgslrelvlagnrlaylqpalf-sglaelreldlsrnalra-ikanvfvqlprl 244
          ::|| | :|::||:|:||:|  ::   | :||: |: | | |: | : :  : |  |: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    245 qklyldrnliaavapgaflglkalrwldl 273
          : | |  | |::::|||| | :|  ::|
          60 RLLSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT   7
ID   R85889 standard; Protein; 603 AA.
AC   R85889;
DT   13-SEP-1996  (first entry)
DE   WD-40 domain-contg. rat insulin-like growth factor binding protein.
KW   WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW   intracellular signalling; protein kinase C; homology; motif; modulator;
KW   receptors of activated protein kinase; enzyme activity; isozyme; human.
OS   Rattus rattus.
PN   WO9521252-A2.
PD   10-AUG-1995.
PF   31-JAN-1995; U01210.
PR   01-FEB-1994; US-190802.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Mochly-Rosen D,  Ron D;
DR   WPI; 95-283772/37.
PT   New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT   activity of a protein, eg. protein kinase C, which interacts with a
PT   protein contg. a WD-40 region.
PS   Example 5; Page 125-128; 351pp; English.
CC   Proteins R85851-92 are protein which contain at least one WD-40 (also
CC   called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC   regions are involved in protein-protein interactions between proteins
CC   involved in intracellular signalling. An example of such an interaction
CC   is between protein kinase C and receptors of activated protein kinase
CC   (RACK), esp. RACK-1 (R85850). Proteins R85851-82 were isolated based on
CC   homology with beta-transducin, whereas proteins R85882-92 were isolated
CC   based on homology with the WD-40 consensus sequence (R85893). The
CC   proteins were used to construct the peptides R84928-R85063 and
CC   R85786-R85842. The peptides can be used to identify target proteins
CC   contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC   proteins involved in protein-protein interaction and to screen for drugs
CC   that will affect protein-protein interaction involving WD-40 domains.
SQ   Sequence  603 AA;

  Query Match          18.0%;  Score 179;  DB 17;  Length 603;
  Best Local Similarity  37.5%;  Pred. No. 5.39e-07;
  Matches   39;  Conservative   22;  Mismatches  41;  Indels   2;  Gaps   2;

Db    356 gafsglfnvavmnlsgnclrslpervfqg-ldklhslhlehsclghvrlhtfaglsglrr 414
          |||:|  :|  : |:|| | ::  | |:| |  | :| |  : :| |   ||||||::|
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    415 lflrdnsissieeqslaglselleldlttnrlthlprqlfqglg 458
          | | || |::|   ::: |  |  ::| :| :     :|  |||
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINLLSNPFNC-NCHLGAGLG 104


RESULT   8
ID   W39170 standard; Protein; 245 AA.
AC   W39170;
DT   08-MAY-1998  (first entry)
DE   Human PKD1 protein fragment 1.
KW   PKD1; polycystin; polycystic renal degeneration; diagnosis; epitope;
KW   therapy.
OS   Homo sapiens.
PN   DE19650758-C1.
PD   02-JAN-1998.
PF   06-DEC-1996; 050758.
PR   06-DEC-1996; DE-050758.
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI   Martens R, Velhagen I, Zentgraf H;
DR   WPI; 98-034057/04.
PT   PKD1 protein fragments, DNA and antibodies - useful for diagnosis
PT   and therapy of polycystic renal degeneration
PS   Claim 3; Page -; 5pp; German.
CC   W39170-W39184 are fragments of the human polycyclin protein, PKD1.
CC   These fragments contain epitopes recognised by PKD1-specific
CC   antibodies and can be used in the detection and diagnosis of the
CC   autosomal dominant condition, polycystic renal degeneration. W39170
CC   corresponds to amino acids 26-270 of the mature PKD1 protein.
CC   Note: This fragment does not appear in the specification but has
CC   been generated from the sequence represented in W23830.
SQ   Sequence  245 AA;

  Query Match          17.6%;  Score 175;  DB 28;  Length 245;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     27 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 84
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db     85 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 143
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    144 i 144
          :
Qy    133 V 133


RESULT   9
ID   W39171 standard; Protein; 455 AA.
AC   W39171;
DT   08-MAY-1998  (first entry)
DE   Human PKD1 protein fragment 2.
KW   PKD1; polycystin; polycystic renal degeneration; diagnosis; epitope;
KW   therapy.
OS   Homo sapiens.
PN   DE19650758-C1.
PD   02-JAN-1998.
PF   06-DEC-1996; 050758.
PR   06-DEC-1996; DE-050758.
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI   Martens R, Velhagen I, Zentgraf H;
DR   WPI; 98-034057/04.
PT   PKD1 protein fragments, DNA and antibodies - useful for diagnosis
PT   and therapy of polycystic renal degeneration
PS   Claim 3; Page -; 5pp; German.
CC   W39170-W39184 are fragments of the human polycyclin protein, PKD1.
CC   These fragments contain epitopes recognised by PKD1-specific
CC   antibodies and can be used in the detection and diagnosis of the
CC   autosomal dominant condition, polycystic renal degeneration. W39171
CC   corresponds to amino acids 26-480 in the mature PKD1 protein.
CC   Note: This fragment does not appear in the specification but has
CC   been generated from the sequence represented in W23830.
SQ   Sequence  455 AA;

  Query Match          17.6%;  Score 175;  DB 28;  Length 455;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
```

```
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     27 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 84
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db     85 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 143
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |. |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    144 i 144
          :
Qy    133 V 133


RESULT   10
ID   W33396 standard; Protein; 4302 AA.
XX
AC   W33396;
XX
DT   01-JUN-1998  (first entry)
DE   Human PKD1 polypeptide.
KW   Human; polycystic kidney disease 1; PKD1; treatment;
KW   autosomal dominant polycystic kidney disease; APKD.
OS   Homo sapiens.
PN   WO9744457-A1.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997; U08799.
XX
PR   03-JUN-1996; US-658136.
PR   24-MAY-1996; US-655360.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI   Qian F;
XX
DR   WPI; 98-018511/02.
DR   N-PSDB; T94012.
XX
PT   Human polycystic kidney disease gene, PKD1 - useful to treat and
PT   diagnose human autosomal or adult onset polycystic kidney disease
XX
PS   Claim 8; Pages 119-138; 257pp; English.
XX
CC   The present sequence is the human polycystic kidney disease 1
CC   (PKD1) polypeptide. The PKD1 cDNA or polypeptide may be used to
CC   treat autosomal dominant polycystic kidney disease (APKD), and
CC   identify carriers of mutant PKD1 genes, i.e. subjects susceptible
CC   to APKD. Antibodies (Ab) that distinguish between normal and mutant
CC   PKD1 sequences can also be used in diagnostic tests. Anti-PKD1 Ab
CC   may also be used to perform subcellular and histochemical
CC   localisation studies, and to block the function of PKD1. Ab are
CC   also useful in rational drug design studies to identify and test
CC   inhibitors of PKD1. Sense and antisense sequences derived from the
CC   PKD1 gene may used for detection and therapy.
XX
SQ   Sequence   4302 AA;

  Query Match          17.6%;  Score 175;  DB 29;  Length 4302;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     52 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 109
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 168
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 i 169
          :
Qy    133 V 133


RESULT   11
ID   W23830 standard; Protein; 4302 AA.
XX
AC   W23830;
XX
DT   08-MAY-1998  (first entry)
DE   Human PKD1 protein.
KW   PKD1; polycystin; polycystic renal degeneration; diagnosis; epitope;
KW   therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Region          26..270
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          26..480
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          361..540
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          480..700
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          541..840
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          700..1100
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          1011..1220
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          2161..2370
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          2723..2931
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          2850..3000
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          2932..3067
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          3100..3280
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          3200..3400
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          3311..3603
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
FT   Region          4090..4302
FT                   /note= "contains epitope recognised by a
FT                         PKD1-specific antibody"
XX
PN   DE19650758-C1.
XX
PD   02-JAN-1998.
XX
PF   06-DEC-1996; 050758.
XX
PR   06-DEC-1996; DE-050758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Martens R, Velhagen I, Zentgraf H;
XX
DR   WPI; 98-034057/04.
XX
PT   PKD1 protein fragments, DNA and antibodies - useful for diagnosis
PT   and therapy of polycystic renal degeneration
XX
PS   Disclosure; Page -; 5pp; German.
XX
CC   This sequence represents a human polycyclin, PKD1. This protein is used
CC   to generate fragments (see W39170-W39184) which contain epitopes
CC   recognised by PKD1-specific antibodies and can be used in the detection
CC   and diagnosis of the autosomal dominant condition, polycystic renal
CC   degeneration.
XX
SQ   Sequence   4302 AA;

  Query Match          17.6%;  Score 175;  DB 28;  Length 4302;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     52 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 109
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 168
```

```
          |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 i 169
          :
Qy    133 V 133


RESULT   12
ID   W00870 standard; Protein; 4302 AA.
AC   W00870;
DT   02-FEB-1997  (first entry)
DE   Polycystic kidney disease 1 (PKD1) polypeptide.
KW   Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;
KW   therapy; polycystin.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..23
FT                   /label= Sig_peptide
FT   protein         24..4303
FT                   /label= Mat_protein
FT   region          72..125
FT                   /label= LRR
FT                   /note= "leucine-rich repeat region"
FT   domain          2580..2600
FT                   /label= TM1
FT                   /note= "transmembrane domain 1"
FT   domain          2693..2713
FT                   /label= TM2
FT                   /note= "transmembrane domain 2"
FT   domain          3075..3095
FT                   /label= TM3
FT                   /note= "transmembrane domain 3"
FT   domain          3281..3301
FT                   /label= TM4
FT                   /note= "transmembrane domain 4"
FT   domain          3323..3343
FT                   /label= TM5
FT                   /note= "transmembrane domain 5"
FT   domain          3559..3579
FT                   /label= TM6
FT                   /note= "transmembrane domain 6"
FT   domain          3582..3612
FT                   /label= TM7
FT                   /note= "transmembrane domain 7"
FT   domain          3669..3689
FT                   /label= TM8
FT                   /note= "transmembrane domain 8"
FT   misc_difference 50
FT                   /label= N-glycosylation_site
FT   misc_difference 89
FT                   /label= N-glycosylation_site
FT   misc_difference 107
FT                   /label= N-glycosylation_site
FT   misc_difference 112
FT                   /label= N-glycosylation_site
FT   misc_difference 187
FT                   /label= N-glycosylation_site
FT   misc_difference 621
FT                   /label= N-glycosylation_site
FT   misc_difference 632
FT                   /label= N-glycosylation_site
FT   misc_difference 746
FT                   /label= N-glycosylation_site
FT   misc_difference 810
FT                   /label= N-glycosylation_site
FT   misc_difference 841
FT                   /label= N-glycosylation_site
FT   misc_difference 854
FT                   /label= N-glycosylation_site
FT   misc_difference 890
FT                   /label= N-glycosylation_site
FT   misc_difference 921
FT                   /label= N-glycosylation_site
FT   misc_difference 1004
FT                   /label= N-glycosylation_site
FT   misc_difference 1034
FT                   /label= N-glycosylation_site
FT   misc_difference 1072
FT                   /label= N-glycosylation_site
FT   misc_difference 1113
FT                   /label= N-glycosylation_site
FT   misc_difference 1178
FT                   /label= N-glycosylation_site
FT   misc_difference 1194
FT                   /label= N-glycosylation_site
FT   misc_difference 1240
FT                   /label= N-glycosylation_site
FT   misc_difference 1269
FT                   /label= N-glycosylation_site
FT   misc_difference 1336
FT                   /label= N-glycosylation_site
FT   misc_difference 1348
FT                   /label= N-glycosylation_site
FT   misc_difference 1382
FT                   /label= N-glycosylation_site
FT   misc_difference 1450
FT                   /label= N-glycosylation_site
FT   misc_difference 1455
FT                   /label= N-glycosylation_site
FT   misc_difference 1474
FT                   /label= N-glycosylation_site
FT   misc_difference 1518
FT                   /label= N-glycosylation_site
FT   misc_difference 1541
FT                   /label= N-glycosylation_site
FT   misc_difference 1554
FT                   /label= N-glycosylation_site
FT   misc_difference 1563
FT                   /label= N-glycosylation_site
FT   misc_difference 1647
FT                   /label= N-glycosylation_site
FT   misc_difference 1661
FT                   /label= N-glycosylation_site
FT   misc_difference 1733
FT                   /label= N-glycosylation_site
FT   misc_difference 1791
FT                   /label= N-glycosylation_site
FT   misc_difference 1834
FT                   /label= N-glycosylation_site
FT   misc_difference 1867
FT                   /label= N-glycosylation_site
FT   misc_difference 1880
FT                   /label= N-glycosylation_site
FT   misc_difference 1991
FT                   /label= N-glycosylation_site
FT   misc_difference 2050
FT                   /label= N-glycosylation_site
FT   misc_difference 2074
FT                   /label= N-glycosylation_site
FT   misc_difference 2125
FT                   /label= N-glycosylation_site
FT   misc_difference 2248
FT                   /label= N-glycosylation_site
FT   misc_difference 2353
FT                   /label= N-glycosylation_site
FT   misc_difference 2395
FT                   /label= N-glycosylation_site
FT   misc_difference 2412
FT                   /label= N-glycosylation_site
FT   misc_difference 2567
FT                   /label= N-glycosylation_site
FT   misc_difference 2578
FT                   /label= N-glycosylation_site
FT   misc_difference 2645
FT                   /label= N-glycosylation_site
```

```
FT   misc_difference 2718
FT                /label= N-glycosylation_site
FT   misc_difference 2754
FT                /label= N-glycosylation_site
FT   misc_difference 2818
FT                /label= N-glycosylation_site
FT   misc_difference 2841
FT                /label= N-glycosylation_site
FT   misc_difference 2878
FT                /label= N-glycosylation_site
FT   misc_difference 2925
FT                /label= N-glycosylation_site
FT   misc_difference 2956
FT                /label= N-glycosylation_site
FT   misc_difference 2994
FT                /label= N-glycosylation_site
FT   misc_difference 3737
FT                /label= N-glycosylation_site
FT   misc_difference 3789
FT                /label= N-glycosylation_site
FT   misc_difference 3844
FT                /label= N-glycosylation_site
PN   WO9534649-A2.
PD   21-DEC-1995.
PF   13-JUN-1995; G01386.
PR   14-JUN-1994; GB-011900.
PR   23-DEC-1994; WO-G02822.
PR   13-APR-1995; GB-007766.
PR   14-APR-1995; US-422582.
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
PI   Breuning MH,  Halley DJJ,  Harris PC,  Hesseling ALW;
PI   Hughes J,  Janssen LAJ, Nellist MD,  Peral B,  Peters DJM;
PI   Roelfsema JH,  Sampson J,  Ward CJ;
DR   WPI; 96-049678/05.
DR   N-PSDB; T13821.
PT   Isolated polycystic kidney disease I gene and its deletion mutants
PT   - useful in diagnosis and treatment of PKD1-associated disease and
PT   in gene therapy
PS   Claim 16; Fig 15; 181pp; English.
CC   PKD1 polypeptide (W00870), designated polycystin, has a role in the
CC   prevention or suppression of adult polycystic kidney disease (APKD).
CC   It is the product of the PKD1 gene (T13821) located on human
CC   chromosome 16.  The gene is mutated in PKD1 patients.  The PKD1 gene
CC   can be used to produce PKD1 polypeptide in transformed host
CC   cells.  The polypeptide is useful in the diagnosis and treatment
CC   of PKD1-associated diseases, and to raise antibodies.
SQ   Sequence   4302 AA;

  Query Match           17.6%;  Score 175;  DB 19;  Length 4302;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches    39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     52 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 109
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 168
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 i 169
          :
Qy    133 V 133


RESULT   13
ID   R75916 standard; Protein; 4339 AA.
AC   R75916;
DT   14-APR-1996  (first entry)
DE   Polycystic kidney disease 1 gene product.
KW   Autosomal dominant polycystic kidney diease; ADPKD;
KW   polycystic kidney disease 1 gene; PKD1; diagnostic; therapy.
OS   Homo sapiens.
PN   WO9518225-A1.
PD   06-JUL-1995.
PF   23-DEC-1994; G02822.
PR   24-DEC-1993; GB-026470.
PR   14-JUN-1994; GB-011900.
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
PI   Breuning MH,  Halley DJJ,  Harris PC,  Hesseling ALW;
PI   Hughes J,  Janssen LAJ, Nellist MD,  Peral B,  Peters DJM;
PI   Roelfsema JH,  Sampson J,  Ward CJ;
DR   WPI; 95-246390/32.
DR   N-PSDB; Q91438.
PT   Isolated poly:cystic kidney disease 1 gene and its mutants - useful
PT   for treatment and diagnosis of autosomal dominant poly:cystic kidney
PT   disease
PS   Disclosure; Fig 10; 119pp; English.
CC   A novel protein (R75916) is encoded by the polycystic kidney disease
CC   1 (PKD1) gene (see Q91438), which maps to 16p13.3.  Mutations at this
CC   locus are associated with autosomal dominant polycystic kidney disease
CC   (ADPKD).  The protein can be used to screen actual or suspected
CC   ADPKD patients for normal or mutated PKD1 polypeptide, or is
CC   used to treat or prevent PKD1-associated disorders such as ADPKD by
CC   administration to affected cells.
SQ   Sequence   4339 AA;

  Query Match           17.6%;  Score 175;  DB 15;  Length 4339;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches    39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db      9 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 66
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db     67 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 125
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    126 i 126
          :
Qy    133 V 133


RESULT   14
ID   R87539 standard; Protein; 4339 AA.
AC   R87539;
DT   02-FEB-1997  (first entry)
DE   Polycystic kidney disease 1 polypeptide (polycystin).
KW   Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;
KW   therapy; polycystin.
OS   Homo sapiens.
PN   WO9534649-A2.
PD   21-DEC-1995.
PF   13-JUN-1995; G01386.
PR   14-JUN-1994; GB-011900.
PR   23-DEC-1994; WO-G02822.
PR   13-APR-1995; GB-007766.
PR   14-APR-1995; US-422582.
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
PI   Breuning MH,  Halley DJJ,  Harris PC,  Hesseling ALW;
PI   Hughes J,  Janssen LAJ, Nellist MD,  Peral B,  Peters DJM;
PI   Roelfsema JH,  Sampson J,  Ward CJ;
DR   WPI; 96-049678/05.
DR   N-PSDB; T08807.
PT   Isolated polycystic kidney disease I gene and its deletion mutants
PT   - useful in diagnosis and treatment of PKD1-associated disease and
```

```
PT   in gene therapy
PS   Claim 18; Fig 10; 181pp; English.
CC   PKD1 polypeptide (R87539) is encoded by a partial cDNA clone
CC   (T08803) corresponding to the complete human PKD1 gene (see also
CC   T13821) apart from its extreme 5' end.  This gene is associated
CC   with adult polycystic kidney disease (APKD).  The polypeptide can
CC   be produced in transformed host cells for use in the diagnosis of
CC   PKD1-associated diseases, for the detection of disease carriers,
CC   and for the treatment or prevention of these diseases.  The PKD1
CC   polypeptide plays a role in the suppression or prevention of APKD.
SQ   Sequence   4339 AA;

  Query Match          17.6%;  Score 175;  DB 19;  Length 4339;
  Best Local Similarity  32.2%;  Pred. No. 1.23e-06;
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db      9 sgrglrtl-gpalrip-adataldvshnllraldvgllanlsalaeldisnnkistleeg 66
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db     67 ifanlfnlseinlsgnpfecdcgl-awlprwaeeqqvrvvqpeaatcagpgslagqpllg 125
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : | |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    126 i 126
          :
Qy    133 V 133


RESULT   15
ID   R90302 standard; Protein; 4303 AA.
AC   R90302;
DT   18-JUL-1996  (first entry)
DE   Polycystic kidney disease 1 gene product.
KW   Polycystic kidney disease; PKD1; autosomal dominant; ADPKD; mutation;
KW   exon; short arm; chromosome 16; repeated region; alternative splicing;
KW   extracellular matrix proteins; antibody; detection; diagnosis;
KW   mini gene therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..23
FT                   /note= "Signal peptide"
FT   peptide         32..65
FT                   /note= "Leucine-rich repeat cystein-rich N-terminus"
FT   domain          72..94
FT                   /note= "Leucine-rich repeat domain, LRR1"
FT   domain          97..119
FT                   /note= "Leucine-rich repeat domain, LRR2"
FT   peptide         123..177
FT                   /note= "Leucine-rich repeat cysteine-rich C-terminus"
FT   domain          282..353
FT                   /note= "PKD domain, R1"
FT   binding_site    405..533
FT                   /note= "C-type lectin binding domain"
FT   domain          639..671
FT                   /note= "LDL module, LDL-A"
FT   domain          1032..1124
FT                   /note= "PKD domain, R2"
FT   domain          1138..1209
FT                   /note= "PKD domain, R3"
FT   domain          1221..1292
FT                   /note= "PKD domain, R4"
FT   domain          1305..1377
FT                   /note= "PKD domain, R5"
FT   domain          1390..1463
FT                   /note= "PKD domain, R6"
FT   domain          1477..1545
FT                   /note= "PKD domain, R7"
FT   domain          1559..1629
FT                   /note= "PKD domain, R8"
FT   domain          1643..1715
FT                   /note= "PKD domain, R9"
FT   domain          1729..1799
FT                   /note= "PKD domain, R10"
FT   domain          1815..1884
FT                   /note= "PKD domain, R11"
FT   domain          1898..1968
FT                   /note= "PKD domain, R12"
FT   domain          1983..2058
FT                   /note= "PKD domain, R13"
FT   domain          2071..2142
FT                   /note= "PKD domain, R14"
PN   WO9534573-A1.
PD   21-DEC-1995.
PF   02-JUN-1995; U07079.
PR   03-JUN-1994; US-253524.
PR   30-MAR-1995; US-413580.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (MILL-) MILLENIUM PHARM INC.
PI   Glucksmann S,  Reeders S,  Schneider M;
DR   WPI; 96-049618/05.
DR   N-PSDB; T11708.
PT   DNA encoding poly-cystic kidney disease gene product - for use in
PT   gene therapy of ADPKD, and in the evaluation of treatment for PKD
PS   Claim 2; Fig 6; 126pp; English.
CC   This sequence represents a polycystic kidney disease (PKD1) product
CC   which is associated with autosomal dominant polycystic kidney
CC   disease (ADPKD).  Mutations within the PKD1 gene are responsible for
CC   approx. 90% of cases of ADPKD.  The coding region of the PKD1 gene
CC   is complex and extensive.  It covers approx. 60 kb and contains a
CC   total of 46 exons.  It has been localised to within a 750 kb chromosomal
CC   region on the short arm of chromosome 16.  Approximately the first two
CC   thirds of the PKD1 gene is duplicated several times in a transcribed
CC   fashion elsewhere in the genome.  The PKD1 gene also contains extensive
CC   repeated regions of high GC content.  A number of the exons have
CC   alternatively spliced forms giving rise to a number of cDNA clones.
CC   The PKD1 protein contains at least 5 distinct peptide domains which
CC   are likely to be involved in protein-protein and/or protein-
CC   carbohydrate interactions.  It also shares amino acid similarity
CC   with a number of extracellular matrix proteins.  Antibodies raised
CC   against the PKD1 protein may be used in the detection of mutant PKD1
CC   and, therefore, diagnosis of ADPKD.  Fragments of the PKD1 gene may be
CC   used in "mini" gene therapy for the treatment of ADPKD.
SQ   Sequence   4303 AA;

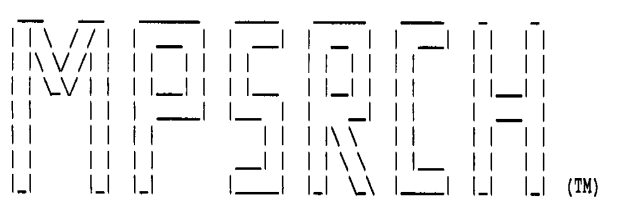
  Query Match          16.8%;  Score 167;  DB 17;  Length 4303;
  Best Local Similarity  32.2%;  Pred. No. 6.42e-06;
  Matches   39;  Conservative   27;  Mismatches  49;  Indels   6;  Gaps   6;

Db     52 sgrglrtl-gpalrip-adateldvshnllraldvgllanlsalaeldisnnkistleeg 109
          :|  | |: | ::|   :    | :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 ifanlfnlseinlsgnpfecdcgl-awlpqwaeeqqvrvvqpeaatcagpgslagqpllg 168
           |: | :|| ||| :|||:|:| | | |  |    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 i 169
          :
Qy    133 V 133


Search completed: Fri May 28 08:47:16 1999
Job time : 67 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:47:53 1999; MasPar time 16.43 Seconds
336.623 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-4
Description: (1-138) from US09191647.pep
Perfect Score: 995
Sequence: 1 EGAFNGAASVQELMLTGNQL.....CQKPFFLKEIPIQGVGHPGI 138

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 42.283; Variance 84.876; scale 0.498

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	326	32.8	1469	2	B36665	slit protein 2 precur	3.15e-40
2	326	32.8	1480	2	A36665	slit protein 1 precur	3.15e-40
3	217	21.8	560	2	A60164	platelet membrane gly	3.68e-20
4	208	20.9	1091	2	A58532	glial cell membrane g	1.40e-18
5	195	19.6	907	2	JE0176	orphan G protein-coup	2.54e-16
6	191	19.2	361	2	A53860	chondroadherin precur	1.24e-15
7	185	18.6	605	2	JC5239	insulin-like growth f	1.30e-14
8	184	18.5	605	2	A41915	insulin-like growth f	1.93e-14
9	183	18.4	536	2	A34901	lysine carboxypeptida	2.85e-14
10	179	18.0	603	2	JC1282	insulin-like growth f	1.35e-13
11	175	17.6	4302	2	A38971	polycystic kidney dis	6.33e-13
12	172	17.3	682	2	A43318	connectin precursor -	2.00e-12
13	172	17.3	682	2	A49121	cell-surface molecule	2.00e-12
14	166	16.7	603	2	JC6128	insulin-like growth f	1.97e-11
15	165	16.6	420	2	A53531	oncofetal trophoblast	2.88e-11
16	162	16.3	313	2	G02020	p37NB - human	8.95e-11
17	158	15.9	1115	2	S40241	G protein-coupled rec	4.01e-10
18	150	15.1	1535	2	S46224	peroxidasin - fruit f	7.73e-09
19	149	15.0	1134	1	A29944	chaoptin precursor -	1.12e-08
20	144	14.5	312	1	NBHUA2	leucine-rich alpha-2-	6.87e-08
21	140	14.1	1025	1	A57676	protein kinase Xa21 (	2.90e-07
22	139	14.0	360	2	S68209	sds22 protein homolog	4.14e-07
23	138	13.9	343	2	A41748	lumican precursor - c	5.91e-07
24	138	13.9	626	1	NBHUIA	platelet glycoprotein	5.91e-07
25	136	13.7	360	2	I47020	decorin - rabbit	1.20e-06
26	136	13.7	661	2	I56258	RP105 - mouse	1.20e-06
27	136	13.7	1097	2	A29943	Toll protein precurso	1.20e-06
28	135	13.6	359	1	NBHUC8	decorin precursor - h	1.71e-06
29	134	13.5	176	1	A46606	platelet glycoprotein	2.43e-06
30	134	13.5	662	2	S42799	garp precursor - huma	2.43e-06
31	133	13.4	230	2	I46918	leucine-rich glycopro	3.45e-06
32	132	13.3	360	2	S06280	decorin precursor - b	4.90e-06
33	132	13.3	684	2	T01267	leucine-rich repeat t	4.90e-06
34	132	13.3	942	1	JQ1674	protein kinase TMK1 (	4.90e-06
35	129	13.0	206	1	NBHUIB	platelet glycoprotein	1.39e-05
36	129	13.0	411	1	I55604	platelet glycoprotein	1.39e-05
37	128	12.9	863	2	A55173	cf-9 protein precurso	1.96e-05
38	127	12.8	690	2	T01183	protein kinase homolo	2.77e-05
39	126	12.7	852	2	I51259	tyrosine kinase C rec	3.90e-05
40	125	12.6	368	1	BGHUN	biglycan precursor -	5.49e-05
41	125	12.6	369	2	S32793	biglycan precursor -	5.49e-05
42	125	12.6	369	2	S20811	proteoglycan I - mous	5.49e-05
43	125	12.6	369	2	S32559	biglycan precursor -	5.49e-05
44	124	12.5	350	2	E71373	probable regulatory p	7.72e-05
45	124	12.5	354	2	A55454	decorin precursor - m	7.72e-05

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
  371-394         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494         #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537         #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695         #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733         #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  846-890         #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061       #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           32.8%;  Score 326;  DB 2;  Length 1469;
  Best Local Similarity  37.5%;  Pred. No. 3.15e-40;
  Matches   45;  Conservative   35;  Mismatches  38;  Indels   2;  Gaps   2;

Db    365 ALSGLKQLTTLVLYGNKIKDLPSGVFKG-LGSLRLLLLNANEISCIRKDAFRDLHSLSLL 423
          |::|   :  |:| || : : :  |:| |::|: |:| :| |:|: :|:|  | |: ||
Qy      3 AFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLL 62

Db    424 SLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRW-LADYLHKNPIETSGARCESP 482
          ||||| | ::: |:| :: |: |::|  ||| |:|:|   |:  |:|  | :: :||: |
Qy     63 SLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing

FEATURE
  66-91         #domain proteoglycan amino-terminal homology #label
                  PAH1\
  101-124       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
  125-148       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
  149-172       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
  173-196       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
  197-220       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
  228-272       #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
  288-313       #domain proteoglycan amino-terminal homology #label
                  PAH2\
  323-346       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
  347-370       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
  371-394       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494       #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537       #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695       #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733       #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  791-814       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
  815-838       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
  846-890       #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061     #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           32.8%;  Score 326;  DB 2;  Length 1480;
  Best Local Similarity  37.5%;  Pred. No. 3.15e-40;
  Matches   45;  Conservative   35;  Mismatches  38;  Indels   2;  Gaps   2;

Db    365 ALSGLKQLTTLVLYGNKIKDLPSGVFKG-LGSLRLLLLNANEISCIRKDAFRDLHSLSLL 423
          |::|   :  |:| || :  : :  |:| |::|: |:| :| |:|: :|:|  | |: ||
Qy      3 AFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLL 62

Db    424 SLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRW-LADYLHKNPIETSGARCESP 482
          ||||| | ::: |:| :: |: |::|  ||| |:|:|   |:  |:|  | :: :||: |
Qy     63 SLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT    3
ENTRY           A60164     #type complete
TITLE           platelet membrane glycoprotein V precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
```

```
DATE            12-Jan-1993 #sequence_revision 24-Feb-1994 #text_change
                  17-Mar-1999
ACCESSIONS      A48030; A60164; A35483; B35483; C35483; A60432; A47507;
                  S34329
REFERENCE       A48030
  #authors      Lanza, F.; Morales, M.; de La Salle, C.; Cazenave, J.P.;
                  Clemetson, K.J.; Shimomura, T.; Phillips, D.R.
  #journal      J. Biol. Chem. (1993) 268:20801-20807
  #title        Cloning and characterization of the gene encoding the human
                  platelet glycoprotein V. A member of the leucine-rich
                  glycoprotein family cleaved during thrombin-induced
                  platelet activation.
  #cross-references MUID:94012616
  #accession    A48030
    ##molecule_type DNA
    ##residues      1-560 ##label LA2
    ##cross-references EMBL:Z23091; NID:g312501; PID:g312502
REFERENCE       A60164
  #authors      Shimomura, T.; Fujimura, K.; Maehama, S.; Takemoto, M.; Oda,
                  K.; Fujimoto, T.; Oyama, R.; Suzuki, M.; Ichihara-Tanaka,
                  K.; Titani, K.; Kuramoto, A.
  #journal      Blood (1990) 75:2349-2356
  #title        Rapid purification and characterization of human platelet
                  glycoprotein V: the amino acid sequence contains
                  leucine-rich repetitive modules as in glycoprotein Ib.
  #cross-references MUID:90275263
  #accession    A60164
    ##molecule_type protein
    ##residues      365-384,'X',386-390,'X',392-395,'X',397;188-208,'I',210;
                    27-50,'X',52-53;174-180,'X',182-187;121-144;145-172;
                    290-297,'X',299-311,'X',313-326,'I';142-151,'X',
                    153-163;'YNTPDRXLAXYGGF';81-105,'XX',108,'T';61-72,
                    'TK',75-77;'V',56-57;'G',479-487,'X',489-498,'X',500,
                    'X',502-503,'X',505,'X',507-508,'D' ##label SHI
REFERENCE       A35483
  #authors      Roth, G.J.; Church, T.A.; McMullen, B.A.; Williams, S.A.
  #journal      Biochem. Biophys. Res. Commun. (1990) 170:153-161
  #title        Human platelet glycoprotein V: a surface leucine-rich
                  glycoprotein related to adhesion.
  #cross-references MUID:90321220
  #accession    A35483
    ##molecule_type protein
    ##residues      145-166,'I',168-169,'X',171-172 ##label ROT
    ##note          this proteolytic fragment was designated peptide M392
  #accession    B35483
    ##molecule_type protein
    ##residues      121-129,'W',131-135;466-468,'X',470 ##label RO2
    ##note          this material was designated peptide M393 but may
                    contain two peptides
  #accession    C35483
    ##molecule_type protein
    ##residues      252-266,'H',268-272,'X',274-279,'I',281-284,'I',286
                    ##label RO3
    ##note          this proteolytic fragment was designated peptide M401
REFERENCE       A60432
  #authors      Zafar, R.S.; Walz, D.A.
  #journal      Thromb. Res. (1989) 53:31-44
  #title        Platelet membrane glycoprotein V: characterization of the
                  thrombin-sensitive glycoprotein from human platelets.
  #cross-references MUID:89162331
  #accession    A60432
    ##molecule_type protein
    ##residues      477-478,'FX',481-485,'E',487,'V',489-492,'NQ',495,'E',
                    497-498 ##label ZAF
REFERENCE       A47507
  #authors      Hickey, M.J.; Hagen, F.S.; Yagi, M.; Roth, G.J.
  #journal      Proc. Natl. Acad. Sci. U.S.A. (1993) 90:8327-8331
  #title        Human platelet glycoprotein V: characterization of the
                  polypeptide and the related Ib-V-IX receptor system of
                  adhesive, leucine-rich glycoproteins.
  #cross-references MUID:93391348
  #accession    A47507
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues      1-560 ##label RES
    ##cross-references GB:L11238; NID:g388759; PID:g388760
COMMENT     This platelet membrane protein is a substrate for thrombin.
COMMENT     The amino end of the intact protein is blocked.
COMMENT     This protein is absent in Bernard-Soulier syndrome.
GENETICS
  #gene         GDB:GP5
    ##cross-references GDB:230236; OMIM:173511
  #map_position 5pter-5qter
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        blocked amino end; glycoprotein; platelet; tandem repeat;
                  transmembrane protein
SUMMARY         #length 560  #molecular-weight 60958  #checksum 7673

  Query Match           21.8%;  Score 217;  DB 2;  Length 560;
  Best Local Similarity  34.4%;  Pred. No. 3.68e-20;
  Matches   43;  Conservative   28;  Mismatches  49;  Indels   5;  Gaps   4;

Db    332 QGAFQGLGELQVLALHSNGLTALPDGLLRG-LGKLRQVSLRRNRLRALPRALFRNLSSLE 390
          :||| | : :| | | :| | ::    :|| |: |: : || | :  ::   |  |||:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    391 SVQLDHNQLETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLG-WLRQHLGLVGGEEPP 449
            : |  |:: |:    | :|  |: : |  |:  |:| ||: || ||| :  :|:|:  |
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRR-IVSGN--P 117

Db    450 RCAGP 454
          ||  |
Qy    118 RCQKP 122


RESULT    4
ENTRY           A58532     #type complete
TITLE           glial cell membrane glycoprotein LIG-1 precursor - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change
                  17-Mar-1999
ACCESSIONS      A58532
REFERENCE       A58532
  #authors      Suzuki, Y.; Sato, N.; Tohyama, M.; Wanaka, A.; Takagi, T.
  #journal      J. Biol. Chem. (1996) 271:22522-22527
  #title        cDNA cloning of a novel membrane glycoprotein that is
                  expressed specifically in glial cells in the mouse brain;
                  LIG-1, a protein with leucine-rich repeats and
                  immunoglobulin-like domains.
  #cross-references MUID:96394313
  #accession    A58532
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues      1-1091 ##label SUZ
    ##cross-references GB:D78572; NID:g1545806; PID:g1545807
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
FEATURE
  36-61             #domain proteoglycan amino-terminal homology #label PAH\
  71-94             #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  95-117            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  118-141           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  142-165           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  166-189           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  191-213           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  214-237           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  238-261           #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LRR8\
  262-285         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  286-309         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  310-333         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  334-357         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  358-381         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  385-408         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  409-432         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  440-485         #domain proteoglycan carboxyl-terminal homology #label
                  PCS1
SUMMARY         #length 1091  #molecular-weight 119283  #checksum 7937

  Query Match          20.9%;  Score 208;  DB 2;  Length 1091;
  Best Local Similarity  33.1%;  Pred. No. 1.40e-18;
  Matches   45;  Conservative   30;  Mismatches  57;  Indels   4;  Gaps   4;

Db    350 EGAFKGLKSLRVLDLDHNEISGTIEDTSGAFTGLDNLSKLTLFGNKIKSVAKRAFSGLES 409
          ||||:|  |:: | |  |::  |:    |   ||  |  | | :| |  |:: :|:|| |
Qy      1 EGAFNGAASVQELMLTGNQLE-TVHG-RGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSS 58

Db    410 LEHLNLGENAIRSVQFDAFAKMKNLKELYISSESFLCDCQLKW-LPPWLMGRMLQAFVTA 468
          :  |:| :| | ::   ||: : :|  : : |::| |:|:|   |  ||  | : :   :
Qy     59 VRLLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVS-GNP 117

Db    469 TCAHPESLKGQSIFSV 484
          |  |  ||  :| :|
Qy    118 RCQKPFFLKEIPIQGV 133


RESULT    5
ENTRY           JE0176     #type complete
TITLE           orphan G protein-coupled receptor precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      JE0176
REFERENCE       JE0176
   #authors     McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.;
                  Caskey, C.T.; Liu, Q.
   #journal     Biochem. Biophys. Res. Commun. (1998) 247:266-270
   #title       Identification and cloning of an orphan G protein-coupled
                  receptor of the glycoprotein hormone receptor subfamily.
   #cross-references MUID:98308104
   #accession   JE0176
      ##molecule_type mRNA
      ##residues      1-907 ##label MCD
      ##cross-references GB:AF062006
COMMENT    This protein is a receptor for a novel class of glycoprotein
             ligands.
GENETICS
   #gene        HG38
   #map_position 12q22-23
FEATURE
   1-21            #domain signal sequence #status predicted #label SIG\
   562-583         #domain transmembrane #status predicted #label TM1\
   594-616         #domain transmembrane #status predicted #label TM2\
   639-660         #domain transmembrane #status predicted #label TM3\
   681-701         #domain transmembrane #status predicted #label TM4\
   725-744         #domain transmembrane #status predicted #label TM5\
   768-791         #domain transmembrane #status predicted #label TM6\
   803-824         #domain transmembrane #status predicted #label TM7
SUMMARY         #length 907  #molecular-weight 99997  #checksum 8790

  Query Match          19.6%;  Score 195;  DB 2;  Length 907;
  Best Local Similarity  40.2%;  Pred. No. 2.54e-16;
  Matches   35;  Conservative   16;  Mismatches  35;  Indels   1;  Gaps   1;

Db    108 GAFTGLYSLKVLMLQNNQLRHVPTEALQN-LRSLQSLRLDANHISYVPPSCFSGLHSLRH 166
          ||| |  |:  |||  |||  |   :::  | :| :| | :| |: |:   |:|| |:|
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    167 LWLDDNALTEIPVQAFRSLSALQAMTL 193
          | | || :| |   || :| :| :: |
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT    6
ENTRY           A53860     #type complete
TITLE           chondroadherin precursor - bovine
ALTERNATE_NAMES  38K leucine-rich protein
ORGANISM        #formal_name Bos primigenius taurus #common_name cattle
DATE            07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change
                  16-Dec-1998
ACCESSIONS      A53860
REFERENCE       A53860
   #authors     Neame, P.J.; Sommarin, Y.; Boynton, R.E.; Heinegard, D.
   #journal     J. Biol. Chem. (1994) 269:21547-21554
   #title       The structure of a 38-kDa leucine-rich protein
                  (chondroadherin) isolated from bovine cartilage.
   #cross-references MUID:94342341
   #accession   A53860
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-361 ##label NEA
      ##cross-references GB:U08018; NID:g470671; PID:g470672
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        disulfide bond
FEATURE
   300-346         #domain proteoglycan carboxyl-terminal homology #label
                     PCH
SUMMARY         #length 361  #molecular-weight 40861  #checksum 369

  Query Match          19.2%;  Score 191;  DB 2;  Length 361;
  Best Local Similarity  30.1%;  Pred. No. 1.24e-15;
  Matches   34;  Conservative   31;  Mismatches  44;  Indels   4;  Gaps   4;

Db    223 VEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDGAFLGVTTLKHVHLENNRL 282
          |:|||:|: | | ::   :|::    | || | :  :   |: :| |::::: : | :||:
Qy     10 VQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRI 69

Db    283 HQL-PSNFP-FDSLETLTLTNNPWKCTCQL-RGLRRWL-EAKTSRPDATCASP 331
          : |: |  : || |: | :|| :| |:|  || :||   :    :: |  |
Qy     70 TTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT    7
ENTRY           JC5239     #type complete
TITLE           insulin-like growth factor acid-labile chain - baboon
ORGANISM        #formal_name Papio sp. #common_name baboon
DATE            17-Apr-1997 #sequence_revision 09-May-1997 #text_change
                  09-May-1997
ACCESSIONS      JC5239
REFERENCE       JC5239
   #authors     Delhanty, P.; Baxter, R.C.
   #journal     Biochem. Biophys. Res. Commun. (1996) 227:897-902
   #title       The cloning and expression of the baboon acid-labile subunit
                  of the insulin-like growth factor binding protein complex.
   #cross-references MUID:97040714
   #contents    liver
   #accession   JC5239
      ##molecule_type mRNA
      ##residues      1-605 ##label DEL
COMMENT    This factor is structurally related to proinsulin and have
             insuline-like metabolic, differentiative, and cell proliferative
             activities.
SUMMARY         #length 605  #molecular-weight 66110  #checksum 1703

  Query Match          18.6%;  Score 185;  DB 2;  Length 605;
```

```
  Best Local Similarity  34.8%;  Pred. No. 1.30e-14;
  Matches   31;  Conservative   28;  Mismatches  27;  Indels   3;  Gaps   3;

Db    187 DAAFRGLGGLRELVLAGNRLAYLQPALF-SGLAELRELDLSRNALRA-IKANVFAQLPRL 244
          ::|| | ::::||:|:||:|  ::   | :||: |: | | |: | : :  : || |: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    245 QKLYLDRNLIAAVAPGAFLGLKALRWLDL 273
          : | |  | |::::|||| | :|  ::|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT    8
ENTRY           A41915     #type complete
TITLE           insulin-like growth factor-binding complex acid-labile chain
                  precursor - human
ALTERNATE_NAMES  Acid-Labile Subunit (ALS)
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
                  20-Mar-1998
ACCESSIONS      A41915
REFERENCE       A41915
   #authors     Leong, S.R.; Baxter, R.C.; Camerato, T.; Dai, J.; Wood, W.I.
   #journal     Mol. Endocrinol. (1992) 6:870-876
   #title       Structure and functional expression of the acid-labile
                  subunit of the insulin-like growth factor-binding protein
                  complex.
   #cross-references MUID:92357025
   #accession   A41915
      ##status       preliminary
      ##molecule_type mRNA; protein
      ##residues     1-605 ##label LEO
      ##cross-references GB:M86826; NID:g184807; PID:g184808
      ##experimental_source liver
      ##note         sequence extracted from NCBI backbone (NCBIP:110171)
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
FEATURE
   75-98            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   99-122           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   123-146          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   147-170          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   171-194          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   195-218          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   219-242          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   243-266          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   267-290          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   291-314          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   315-338          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   339-362          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12\
   363-386          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR13\
   387-410          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR14\
   411-434          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR15\
   435-458          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR16\
   459-482          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR17\
   483-506          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR18\
   507-529          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR19
SUMMARY         #length 605  #molecular-weight 66034  #checksum 1870

  Query Match           18.5%;  Score 184;  DB 2;  Length 605;
  Best Local Similarity  34.8%;  Pred. No. 1.93e-14;
  Matches   31;  Conservative   27;  Mismatches  28;  Indels   3;  Gaps   3;

Db    187 DAAFRGLGSLRELVLAGNRLAYLQPALF-SGLAELRELDLSRNALRA-IKANVFVQLPRL 244
          ::|| | :|::||:|:||:|  ::   | :||: |: | | |: | : :  : |  |: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    245 QKLYLDRNLIAAVAPGAFLGLKALRWLDL 273
          : | |  | |::::|||| | :|  ::|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT    9
ENTRY           A34901     #type complete
TITLE           lysine carboxypeptidase (EC 3.4.17.3) 83K chain - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change
                  24-Sep-1998
ACCESSIONS      A34901
REFERENCE       A34901
   #authors     Tan, F.; Weerasinghe, D.K.; Skidgel, R.A.; Tamei, H.; Kaul,
                  R.K.; Roninson, I.B.; Schilling, J.W.; Erdoes, E.G.
   #journal     J. Biol. Chem. (1990) 265:13-19
   #title       The deduced protein sequence of the human carboxypeptidase N
                  high molecular weight subunit reveals the presence of
                  leucine-rich tandem repeats.
   #cross-references MUID:90094386
   #accession   A34901
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-536 ##label TAN
      ##cross-references GB:J05158; NID:g179935; PID:g179936
GENETICS
   #gene        GDB:ACBP
      ##cross-references GDB:127893
   #map_position 6q25.3-6q26
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        hydrolase; metallo-carboxypeptidase
FEATURE
   77-100           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   101-124          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   125-148          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   149-172          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   173-196          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   197-220          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   221-244          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   245-268          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   269-292          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   293-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   317-340          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   341-364          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12
SUMMARY         #length 536  #molecular-weight 58649  #checksum 8569

  Query Match           18.4%;  Score 183;  DB 2;  Length 536;
```

```
  Best Local Similarity  37.4%;  Pred. No. 2.85e-14;
  Matches   34;  Conservative   22;  Mismatches  34;  Indels   1;  Gaps   1;

Db    117 EGLFQHLAALESLHLQGNQLQALPRRLFQP-LTHLKTLNLAQNLLAQLPEELFHPLTSLQ 175
          || |   |::: | | ||||:::  | |:  |: |||| |  ||:: :::: |  |:|::
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    176 TLKLSNNALSGLPQGVFGKLGSLQELFLDSN 206
          | | :| :: :  | |  | ||  : | ||
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSN 91


RESULT   10
ENTRY            JC1282     #type complete
TITLE            insulin-like growth factor-binding protein acid labile chain
                   precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                   15-Aug-1997
ACCESSIONS       JC1282
REFERENCE        JC1282
   #authors      Dai, J.; Baxter, R.C.
   #journal      Biochem. Biophys. Res. Commun. (1992) 188:304-309
   #title        Molecular cloning of the acid-labile subunit of the rat
                   insulin-like growth factor binding protein complex.
   #cross-references MUID:93038676
   #accession    JC1282
      ##molecule_type mRNA
      ##residues      1-603 ##label DAI
      ##experimental_source liver
      ##note          the authors translated the codon AAG for residue 63 as
                       Arg, AAA for residue 205 as Pro and GGT for residue
                       260 as Arg
CLASSIFICATION   #superfamily leucine-rich alpha-2-glycoprotein repeat
                   homology
FEATURE
   1-27              #domain signal sequence #status predicted #label SIG\
   28-603            #product insulin-like growth factor binding protein,
                       acid labile chain #status predicted #label MAT
SUMMARY          #length 603  #molecular-weight 66811  #checksum 8075

  Query Match           18.0%;  Score 179;  DB 2;  Length 603;
  Best Local Similarity  37.5%;  Pred. No. 1.35e-13;
  Matches   39;  Conservative   22;  Mismatches  41;  Indels   2;  Gaps   2;

Db    356 GAFSGLFNVAVMNLSGNCLRSLPERVFQG-LDKLHSLHLEHSCLGHVRLHTFAGLSGLRR 414
          |||:|  :|  : |:|| | ::  | |:| |  | :| |  : :| |   ||||||::|
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

      415 LFLRDNSISSIEEQSLAGLSELLELDLTTNRLTHLPRQLFQGLG 458
          | | || |::|   ::: |  |  ::| :| :     :|  |||
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINLLSNPFNC-NCHLGAGLG 104


RESULT   11
ENTRY            A38971     #type complete
TITLE            polycystic kidney disease protein 1 precursor - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change
                   16-Dec-1998
ACCESSIONS       A38971; A56520; A44604
REFERENCE        A38971
   #authors      Harris, P.C.
   #submission   submitted to GenBank, May 1995
   #accession    A38971
      ##molecule_type mRNA
      ##residues      1-4302 ##label HAR
      ##cross-references GB:L33243; NID:g904222; PID:g904223
REFERENCE        A56520
   #authors      Alexandra Gluecksmann-Kuis, M.; Tayber, O.; Woolf, E.A.;
                   Bougueleret, L.; Deng, N.; Alperin, G.D.; Iris, F.;
                   Hawkins, F.; Munro, C.; Lakey, N.; Duyk, G.; Schneider,
                   M.C.; Geng, L.; Zhang, F.; Zhao, Z.; Torosian, S.; Zhou,
                   J.; Reeders, S.T.; Bork, P.; Pohischmidt, M.; Loehning, C.;
                   Kraus, B.; Nowicka, U.; Leung, A.L.S.; Frischauf, A.M.
   #journal      Cell (1995) 81:289-298
   #title        Polycystic kidney disease: the complete structure of the PKD1
                   gene and its protein.
   #accession    A56520
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-70,'E',72-137,'Q',139-252,'A',254-301,'D',303-690,'P',
                       692-738,'R',740-762,'G',764-773,'QR',776-791,'L',
                       793-865,'L',867-883,'A',885-1055,'T',1057-1091,'T',
                       1093-1276,'G',1278-1723,'A',1725-1975,'V',1977-3389,
                       'Q',3390-3980,'HV',3983-4003,'HV',4006-4302 ##label
                       ALE
      ##cross-references GB:U24497; NID:g799334; PID:g799335; GB:U24499
REFERENCE        A44604
   #authors      Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.;
                   MacCarthy, A.B.; Sloane-Stanley, J.; Buckle, V.J.; Kearney,
                   L.; Higgs, D.R.; Ratcliffe, P.J.; Harris, P.C.; Roelfsema,
                   J.H.; Spruit, L.; Saris, J.J.; Dauwerse, H.G.; Peters,
                   D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.T.;
                   Maheshwar, M.M.; Cordeiro, I.; Santos, H.; Cabral, P.;
                   Sampson, J.R.; Janssen, B.; Hesseling-Janssen, A.L.W.; van
                   den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout,
                   D.; Halley, D.J.J.
   #journal      Cell (1994) 77:881-894
   #title        The polycystic kidney disease 1 gene encodes a 14 kb
                   transcript and lies within a duplicated region on
                   chromosome 16.
   #accession    A44604
      ##status       significant sequence differences
      ##molecule_type mRNA
      ##cross-references GB:L33243
REFERENCE        A38972
   #authors      Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.;
                   MacCarthy, A.B.; Sloane-Stanley, J.; Buckle, V.J.; Kearney,
                   L.; Higgs, D.R.; Ratcliffe, P.J.; Harris, P.C.; Roelfsema,
                   J.H.; Spruit, L.; Saris, J.J.; Dauwerse, H.G.; Peters,
                   D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.T.;
                   Maheshwar, M.M.; Cordeiro, I.; Santos, H.; Cabral, P.;
                   Sampson, J.R.; Janssen, B.; Hesseling-Janssen, A.L.W.; van
                   den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout,
                   D.; Halley, D.J.J.
   #journal      Cell (1994) 78:725
   #contents     annotation; erratum
   #note         this is a revision to the sequence from reference A44604
REFERENCE        A56732
   #authors      Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.;
                   MacCarthy, A.B.; Sloane-Stanley, J.; Buckle, V.J.; Kearney,
                   L.; Higgs, D.R.; Ratcliffe, P.J.; Harris, P.C.; Roelfsema,
                   J.H.; Spruit, L.; Saris, J.J.; Dauwerse, H.G.; Peters,
                   D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.T.;
                   Maheshwar, M.M.; Cordeiro, I.; Santos, H.; Cabral, P.;
                   Sampson, J.R.; Janssen, B.; Hesseling-Janssen, A.L.W.; van
                   den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout,
                   D.; Halley, D.J.J.
   #journal      Cell (1995) 81:1171
   #contents     annotation; erratum
   #note         this is a revision to the sequence from reference A44604
GENETICS
   #gene         GDB:PKD1
      ##cross-references GDB:120293; OMIM:173900; OMIM:601313
   #map_position 16p13.3-16p13.3
CLASSIFICATION   #superfamily proteoglycan carboxyl-terminal homology
KEYWORDS         duplication
FEATURE
   1-23              #domain signal sequence #status predicted #label SIG\
   123-170           #domain proteoglycan carboxyl-terminal homology #label
                       PCS1
SUMMARY          #length 4302  #molecular-weight 462436  #checksum 2698

  Query Match           17.6%;  Score 175;  DB 2;  Length 4302;
  Best Local Similarity  32.2%;  Pred. No. 6.33e-13;
```

```
  Matches   39;  Conservative   29;  Mismatches  47;  Indels   6;  Gaps   6;

Db     52 SGRGLRTL-GPALRIP-ADATALDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEG 109
          :|  | |: | ::|   :   :| :  ||:  :    :| ||::  | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 IFANLFNLSEINLSGNPFECDCGL-AWLPRWAEEQQVRVVQPEAATCAGPGSLAGQPLLG 168
           |: | :|| ||| :|||:|:| | | | :|    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 I 169
          :
Qy    133 V 133


RESULT   12
ENTRY           A43318     #type complete
TLE             connectin precursor - fruit fly (Drosophila melanogaster)
GANISM          #formal_name Drosophila melanogaster
DATE            31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
                  24-Sep-1998
ACCESSIONS      A43318; S28464
REFERENCE       A43318
   #authors     Nose, A.; Mahajan, V.B.; Goodman, C.S.
   #journal     Cell (1992) 70:553-567
   #title       Connectin: a homophilic cell adhesion molecule expressed on a
                  subset of muscles and the motoneurons that innervate them
                  in Drosophila.
   #cross-references MUID:92370678
   #accession   A43318
      ##molecule_type mRNA
      ##residues      1-682 ##label NOS
      ##cross-references GB:M96647; NID:g157083; PID:g157084
      ##note          sequence extracted from NCBI backbone (NCBIN:111422,
                       NCBIP:111423)
REFERENCE       S28464
   #authors     Gould, A.P.; White, R.A.H.
   #submission  submitted to the EMBL Data Library, October 1992
   #description Connectin a target of homeotic gene control in drosophila.
   #accession   S28464
      ##molecule_type mRNA
      ##residues      1-630,'G',632-673,675-678,'M',679-682 ##label GOU
      ##cross-references EMBL:X68701; NID:g7737; PID:g7738
GENETICS
   #gene        FlyBase:Con
      ##cross-references FlyBase:FBgn0005775
SSIFICATION     #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
FEATURE
   1-26             #domain signal sequence #status predicted #label SIG\
   27-682           #product connectin #status predicted #label MAT
SUMMARY         #length 682  #molecular-weight 75991  #checksum 7269

  Query Match          17.3%;  Score 172;  DB 2;  Length 682;
  Best Local Similarity  39.1%;  Pred. No. 2.00e-12;
  Matches   36;  Conservative   19;  Mismatches  34;  Indels   3;  Gaps   3;

Db    263 EGLFADMARLTFLNLAHNQINVLTSEIFRG-LGNLNVLKLTRNNL-NFIGDTVFAELWSL 320
          || |   | :  | |: ||:: : :  ||| |: |: |:| |:||   :::  || | |:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    321 SELELDDNRIERISERALDGLNTLKTLNLRNN 352
            | | ||||  |:  |:  | :| |:|| :|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINLLSN 91


RESULT   13
ENTRY           A49121     #type complete
TITLE           cell-surface molecule connectin - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  20-Mar-1998
ACCESSIONS      A49121
REFERENCE       A49121
   #authors     Gould, A.P.; White, R.A.
   #journal     Development (1992) 116:1163-1174
   #title       Connectin, a target of homeotic gene control in Drosophila.
   #cross-references MUID:93202002
   #accession   A49121
      ##status       preliminary
      ##molecule_type nucleic acid
      ##residues     1-682 ##label GOU
      ##cross-references GB:X68701; NID:g7737; PID:g7738
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIN:127661,
                      NCBIP:127664)
GENETICS
   #gene        FlyBase:Con
      ##cross-references FlyBase:FBgn0005775
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
FEATURE
   199-222          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   223-246          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   247-270          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   271-294          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   295-318          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   319-342          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   343-366          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7
SUMMARY         #length 682  #molecular-weight 75922  #checksum 7093

  Query Match          17.3%;  Score 172;  DB 2;  Length 682;
  Best Local Similarity  39.1%;  Pred. No. 2.00e-12;
  Matches   36;  Conservative   19;  Mismatches  34;  Indels   3;  Gaps   3;

Db    263 EGLFADMARLTFLNLAHNQINVLTSEIFRG-LGNLNVLKLTRNNL-NFIGDTVFAELWSL 320
          || |   | :  | |: ||:: : :  ||| |: |: |:| |:||   :::  || | |:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    321 SELELDDNRIERISERALDGLNTLKTLNLRNN 352
            | | ||||  |:  |:  | :| |:|| :|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINLLSN 91


RESULT   14
ENTRY           JC6128     #type complete
TITLE           insulin-like growth factor binding complex acid labile chain
                  - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            23-Mar-1997 #sequence_revision 09-May-1997 #text_change
                  10-Sep-1997
ACCESSIONS      JC6128
REFERENCE       JC6128
   #authors     Boisclair, Y.R.; Seto, D.; Hsieh, S.; Hurst, K.R.; Ooi, G.T.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1996) 93:10028-10033
   #title       Organization and chromosomal localization of the gene
                  encoding the mouse acid labile subunit of the insulin-like
                  growth factor binding complex.
   #cross-references MUID:96413591
   #accession   JC6128
      ##molecule_type DNA
      ##residues     1-603 ##label BOI
      ##cross-references GB:U66900; NID:g1621612; PID:g1621613
COMMENT    This protein is a serum protein and it is of the ternary complex in
             the physiology of circulating insulin-like growth factor.
GENETICS
   #gene        als
   #map_position 17
```

```
SUMMARY         #length 603  #molecular-weight 66959  #checksum 7670

  Query Match           16.7%;  Score 166;  DB 2;  Length 603;
  Best Local Similarity  32.1%;  Pred. No. 1.97e-11;
  Matches    34;  Conservative   27;  Mismatches  43;  Indels   2;  Gaps   2;

Db    283 EDTFPGLLGLHVLRLAHNAITSLRPRTFKD-LHFLEELQLGHNRIRQLGEKTFEGLGQLE 341
          | :| |  ::: | |: | : ::: | |:  |  |  | |  | |  ::: || ||: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    342 VLTLNDNQIHEVKVGAFFGLFNVAVMNLSGNCLRSLPEHVFQGLGR 387
          :|:| ||:|  :  |||  | ::: :|| :| :     |:  |||:
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNC-HLGAGLGK 105


RESULT   15
ENTRY           A53531     #type complete
TITLE           oncofetal trophoblast glycoprotein 5T4 precursor - human
ALTERNATE_NAMES oncofetal antigen 5T4 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change
                  10-Sep-1997
ACCESSIONS      A53531; S40087
REFERENCE       A53531
   #authors     Myers, K.A.; Rahi-Saund, V.; Davison, M.D.; Young, J.A.;
                  Cheater, A.J.; Stern, P.L.
   #journal     J. Biol. Chem. (1994) 269:9319-9324
   #title       Isolation of a cDNA encoding 5T4 oncofetal trophoblast
                  glycoprotein. An antigen associated with metastasis
                  contains leucine-rich repeats.
   #cross-references MUID:94179356
   #accession   A53531
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-420 ##label MEY
      ##cross-references EMBL:Z29083; NID:g435654; PID:g435655
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        duplication; glycoprotein; transmembrane protein
FEATURE
   1-31             #domain signal sequence #status predicted #label SIG\
   32-420           #product oncofetal trophoblast glycoprotein 5T4 #status
                      predicted #label MAT
SUMMARY         #length 420  #molecular-weight 46031  #checksum 8580

  Query Match           16.6%;  Score 165;  DB 2;  Length 420;
  Best Local Similarity  28.7%;  Pred. No. 2.88e-11;
  Matches    31;  Conservative   31;  Mismatches  42;  Indels   4;  Gaps   4;

Db    215 LELASNHFLYLP-RDVLAQLPSLRHLDLSNNSLVSLTYVSFRNLTHLESLHLEDNALKVL 273
          | |::|::  :  |   : |::|: | | :| :  ::  :|  |: :  | | || :  :
Qy     13 LMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTI 72

Db    274 HNGTLAELQGLPHIRVFLDNNPWVCDCHM-ADMVTWLKETEVVQGKDR 320
            |::: | :|: | : | :||  |:||: | :  ||:   :| |: |
Qy     73 TPGAFTTLVSLSTINL-L-SNPFNCNCHLGAGLGKWLRKRRIVSGNPR 118


Search completed: Fri May 28 08:48:45 1999
Job time : 52 secs.
```

MASPAR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:49:03 1999; MasPar time 5.87 Seconds
665.018 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-4
Description: (1-138) from US09191647.pep
Perfect Score: 995
Sequence: 1 EGAFNGAASVQELMLTGNQL.....CQKPFFLKEIPIQGVGHPGI 138

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 43.452; Variance 74.748; scale 0.581

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	326	32.8	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	5.61e-47
2	217	21.8	560	1	GPV_HUMAN	PLATELET GLYCOPROTEIN	1.03e-23
3	191	19.2	361	1	CHAD_BOVIN	CHONDROADHERIN PRECURS	1.76e-18
4	185	18.6	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	2.67e-17
5	184	18.5	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	4.20e-17
6	183	18.4	536	1	CBP8_HUMAN	CARBOXYPEPTIDASE N 83	6.59e-17
7	179	18.0	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	3.97e-16
8	176	17.7	567	1	GPV_RAT	PLATELET GLYCOPROTEIN	1.51e-15
9	172	17.3	682	1	CONN_DROME	CONNECTIN PRECURSOR.	8.93e-15
10	171	17.2	567	1	GPV_MOUSE	PLATELET GLYCOPROTEIN	1.39e-14
11	167	16.8	4303	1	PKD1_HUMAN	POLYCYSTIN PRECURSOR (	8.06e-14
12	166	16.7	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	1.25e-13
13	158	15.9	1115	1	GPCR_LYMST	G-PROTEIN COUPLED RECE	4.01e-12
14	149	15.0	1134	1	CHAO_DROME	CHAOPTIN PRECURSOR (PH	1.85e-10
15	144	14.5	312	1	A2GL_HUMAN	LEUCINE-RICH ALPHA-2-G	1.50e-09
16	138	13.9	343	1	LUM_CHICK	LUMICAN PRECURSOR (LUM	1.77e-08
17	138	13.9	626	1	GPBA_HUMAN	PLATELET GLYCOPROTEIN	1.77e-08
18	136	13.7	1097	1	TOLL_DROME	TOLL PROTEIN PRECURSOR	4.00e-08
19	135	13.6	359	1	PGS2_HUMAN	BONE PROTEOGLYCAN II P	6.00e-08
20	134	13.5	177	1	GPIX_HUMAN	PLATELET GLYCOPROTEIN	8.99e-08
21	134	13.5	662	1	GARP_HUMAN	GARP PROTEIN PRECURSOR	8.99e-08
22	132	13.3	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	2.01e-07
23	132	13.3	942	1	TMK1_ARATH	PUTATIVE RECEPTOR PROT	2.01e-07
24	130	13.1	360	1	PGS2_CANFA	BONE PROTEOGLYCAN II P	4.47e-07
25	129	13.0	206	1	GPBB_HUMAN	PLATELET GLYCOPROTEIN	6.65e-07
26	126	12.7	852	1	TRKC_CHICK	NT-3 GROWTH FACTOR REC	2.17e-06
27	125	12.6	368	1	PGS1_HUMAN	BONE/CARTILAGE PROTEOG	3.22e-06
28	125	12.6	369	1	PGS1_BOVIN	BONE/CARTILAGE PROTEOG	3.22e-06
29	125	12.6	369	1	PGS1_MOUSE	BONE/CARTILAGE PROTEOG	3.22e-06
30	125	12.6	369	1	PGS1_RAT	BONE/CARTILAGE PROTEOG	3.22e-06
31	124	12.5	354	1	PGS2_MOUSE	BONE PROTEOGLYCAN II P	4.75e-06
32	124	12.5	354	1	PGS2_RAT	BONE PROTEOGLYCAN II P	4.75e-06
33	123	12.4	206	1	GPBB_MOUSE	PLATELET GLYCOPROTEIN	7.01e-06
34	123	12.4	369	1	PGS1_CANFA	BONE/CARTILAGE PROTEOG	7.01e-06
35	122	12.3	818	1	TRKB_CHICK	BDNF / NT-3 GROWTH FAC	1.03e-05
36	121	12.2	839	1	TRKC_HUMAN	NT-3 GROWTH FACTOR REC	1.52e-05
37	120	12.1	825	1	TRKC_PIG	NT-3 GROWTH FACTOR REC	2.23e-05
38	119	12.0	357	1	PGS2_CHICK	BONE PROTEOGLYCAN II P	3.27e-05
39	118	11.9	796	1	TRKA_HUMAN	HIGH AFFINITY NERVE GR	4.78e-05
40	118	11.9	821	1	TRKB_RAT	BDNF / NT-3 GROWTH FAC	4.78e-05
41	118	11.9	821	1	TRKB_MOUSE	BDNF / NT-3 GROWTH FAC	4.78e-05
42	117	11.8	440	1	OMGP_MOUSE	OLIGODENDROCYTE-MYELIN	6.99e-05
43	115	11.6	208	1	GPBB_PAPCY	PLATELET GLYCOPROTEIN	1.48e-04
44	115	11.6	864	1	TRKC_RAT	NT-3 GROWTH FACTOR REC	1.48e-04
45	114	11.5	440	1	OMGP_HUMAN	OLIGODENDROCYTE-MYELIN	2.16e-04

ALIGNMENTS

```
RESULT   1
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          32.8%;  Score 326;  DB 1;  Length 1480;
  Best Local Similarity  37.5%;  Pred. No. 5.61e-47;
  Matches   45;  Conservative   35;  Mismatches  38;  Indels   2;  Gaps   2;

Db    365 ALSGLKQLTTLVLYGNKIKDLPSGVFKG-LGSLRLLLLNANEISCIRKDAFRDLHSLSLL 423
          |::|   :  |:| || :  : :  |:| |::|: |:| :| |:|: :|:|  | |: ||
Qy      3 AFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLL 62

Db    424 SLYDNNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRW-LADYLHKNPIETSGARCESP 482
          ||||| | ::: |:|·:: |: |::|  ||| |:|:|   |:  |:|  | :: :||: |
Qy     63 SLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT   2
ID   GPV_HUMAN      STANDARD;      PRT;   560 AA.
AC   P40197;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 93391348.
RA   HICKEY M.J., HAGEN F.S., YAGI M., ROTH G.J.;
RT   "Human platelet glycoprotein V: characterization of the polypeptide
RT   and the related Ib-V-IX receptor system of adhesive, leucine-rich
RT   glycoproteins.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:8327-8331(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLATELET;
RX   MEDLINE; 94012616.
RA   LANZA F., MORALES M., DE LA SALLE C., CAZENAVE J.-P., CLEMETSON K.J.,
RA   SHIMOMURA T., PHILLIPS D.R.;
RT   "Cloning and characterization of the gene encoding the human platelet
RT   glycoprotein V. A member of the leucine-rich glycoprotein family
RT   cleaved during thrombin-induced platelet activation.";
RL   J. BIOL. CHEM. 268:20801-20807(1993).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   TISSUE=PLATELET;
RX   MEDLINE; 90275263.
RA   SHIMOMURA T., FUJIMURA K., MAEHAMA S., TAKEMOTO M., ODA K.,
RA   FUJIMOTO T., OYAMA R., SUZUKI M., ICIHARA-TANAKA K., TITANI K.,
RA   KURAMOTO A.;
```

```
RT   "Rapid purification and characterization of human platelet
RT   glycoprotein V: the amino acid sequence contains leucine-rich
RT   repetitive modules as in glycoprotein Ib.";
RL   BLOOD 75:2349-2356(1990).
RN   [4]
RP   PARTIAL SEQUENCE.
RC   TISSUE=PLATELET;
RX   MEDLINE; 90321220.
RA   ROTH G.J., CHURCH T.A., MCMULLEN B.A., WILLIAMS S.A.;
RT   "Human platelet glycoprotein V: a surface leucine-rich glycoprotein
RT   related to adhesion.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 170:153-161(1990).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: PLATELETS AND MEGAKARYOCYTES.
CC   -!- PTM: THE N-TERMINAL IS BLOCKED.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11238; G388760; -.
DR   EMBL; Z23091; G312502; -.
DR   MIM; 173511; -.
DR   PFAM; PF00560; LRR; 8.
DR   HSSP; P16473; 1XUM.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   560       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17   523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   524   544       POTENTIAL.
FT   DOMAIN     545   560       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55   415       LEUCINE-RICH REPEATS.
FT   REPEAT      55    78       LRR 1.
FT   REPEAT      79   102       LRR 2.
FT   REPEAT     103   126       LRR 3.
FT   REPEAT     127   150       LRR 4.
FT   REPEAT     151   174       LRR 5.
FT   REPEAT     175   198       LRR 6.
FT   REPEAT     199   222       LRR 7.
FT   REPEAT     223   246       LRR 8.
FT   REPEAT     247   270       LRR 9.
FT   REPEAT     271   294       LRR 10.
FT   REPEAT     295   318       LRR 11.
FT   REPEAT     319   343       LRR 12.
FT   REPEAT     346   367       LRR 13.
FT   REPEAT     368   391       LRR 14.
FT   REPEAT     392   415       LRR 15.
FT   CARBOHYD    51    51
FT   CARBOHYD   181   181
FT   CARBOHYD   243   243       POTENTIAL.
FT   CARBOHYD   267   267       POTENTIAL.
FT   CARBOHYD   298   298
FT   CARBOHYD   312   312
FT   CARBOHYD   385   385
FT   CARBOHYD   499   499       POTENTIAL.
FT   CONFLICT    73    74       MT -> TK (IN REF. 2).
FT   CONFLICT   109   109       K -> T (IN REF. 2).
FT   CONFLICT   130   130       D -> W (IN REF. 3).
FT   CONFLICT   136   138       GID -> PGG (IN REF. 3).
FT   CONFLICT   209   209       L -> I (IN REF. 2).
FT   CONFLICT   267   267       N -> H (IN REF. 3).
FT   CONFLICT   327   327       L -> I (IN REF. 2).
FT   CONFLICT   478   478       P -> G (IN REF. 2).
FT   CONFLICT   509   509       P -> D (IN REF. 2).
SQ   SEQUENCE   560 AA;  60959 MW;  FD65EDD2 CRC32;

  Query Match           21.8%;  Score 217;  DB 1;  Length 560;
  Best Local Similarity  34.4%;  Pred. No. 1.03e-23;
  Matches    43;  Conservative   28;  Mismatches  49;  Indels   5;  Gaps   4;

Db    332 QGAFQGLGELQVLALHSNGLTALPDGLLRG-LGKLRQVSLRRNRLRALPRALFRNLSSLE 390
          :||| | : :| | | :| | ::    :|| |: |: : || | :  ::   |  |||:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    391 SVQLDHNQLETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLG-WLRQHLGLVGGEEPP 449
            : |  |:: |:    | :|  |: : |  |:  |:| ||: || ||| :  :|:|:  |
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRR-IVSGN--P 117

Db    450 RCAGP 454
          ||  |
Qy    118 RCQKP 122


RESULT   3
ID   CHAD_BOVIN     STANDARD;      PRT;   361 AA.
AC   Q27972;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CHONDROADHERIN PRECURSOR (CARTILAGE LEUCINE-RICH PROTEIN) (38 KD BONE
DE   PROTEIN).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 94342341.
RA   NEAME P.J., SOMMARIN Y., BOYNTON R.E., HEINEGARD D.;
RT   "The structure of a 38-kDa leucine-rich protein (chondroadherin)
RT   isolated from bovine cartilage.";
RL   J. BIOL. CHEM. 269:21547-21554(1994).
RN   [2]
RP   SEQUENCE OF 25-55 AND 77-97.
RC   TISSUE=BONE;
RX   MEDLINE; 95113864.
RA   HU B., COULSON L., MOYER B., PRICE P.A.;
RT   "Isolation and molecular cloning of a novel bone phosphoprotein
RT   related in sequence to the cystatin family of thiol protease
RT   inhibitors.";
RL   J. BIOL. CHEM. 270:431-436(1995).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08018; G470672; -.
DR   PFAM; PF00560; LRR; 5.
KW   REPEAT; SIGNAL.
FT   SIGNAL       1    24       OR 23 (IN SOME ISOFORM(S)).
FT   CHAIN       25   361       CHONDROADHERIN.
FT   CHAIN       25   352       CHONDROADHERIN, MINOR FORM.
FT   DOMAIN      79   317       10 X 24 AA LEUCINE-RICH TANDEM REPEATS.
FT   REPEAT      79   102       1.
FT   REPEAT     103   126       2.
FT   REPEAT     127   150       3.
FT   REPEAT     151   174       4.
FT   REPEAT     175   198       5.
FT   REPEAT     199   222       6.
FT   REPEAT     223   246       7.
```

```
FT   REPEAT      248    271       8.
FT   REPEAT      272    293       9.
FT   REPEAT      294    317       10.
FT   DISULFID    306    348
FT   DISULFID    308    328
FT   CONFLICT     25     25       C -> Y (IN REF. 2).
FT   CONFLICT     29     29       C -> W (IN REF. 2).
FT   CONFLICT     31     31       C -> H (IN REF. 2).
FT   CONFLICT     40     40       C -> L (IN REF. 2).
FT   CONFLICT     52     52       S -> R (IN REF. 2).
SQ   SEQUENCE   361 AA;  40884 MW;  A370BB91 CRC32;

  Query Match           19.2%;  Score 191;  DB 1;  Length 361;
  Best Local Similarity  30.1%;  Pred. No. 1.76e-18;
  Matches   34;  Conservative   31;  Mismatches  44;  Indels   4;  Gaps   4;

Db    223 VEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDGAFLGVTTLKHVHLENNRL 282
          |:||:|: | | ::   :|::    | || | :  :   |: :| |::::: : | :||:
Qy     10 VQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRI 69

Db    283 HQL-PSNFP-FDSLETLTLTNNPWKCTCQL-RGLRRWL-EAKTSRPDATCASP 331
           : |: |  : || |: | :|| :| |:|  || :||   :    :: |  |
Qy     70 TTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT    4
ID   ALS_PAPPA      STANDARD;      PRT;   605 AA.
AC   002833;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   PAPIO PAPIO (GUINEA BABOON).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97040714.
RA   DELHANTY P., BAXTER R.C.;
RT   "The cloning and expression of the baboon acid-labile subunit of the
RT   insulin-like growth factor binding protein complex.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 227:897-902(1996).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83462; E323796; -.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN       79    536       LEUCINE-RICH REPEATS.
FT   REPEAT       79     89       LRR 1.
FT   REPEAT       90    113       LRR 2.
FT   REPEAT      114    137       LRR 3.
FT   REPEAT      138    161       LRR 4.
FT   REPEAT      162    185       LRR 5.
FT   REPEAT      186    209       LRR 6.
FT   REPEAT      210    233       LRR 7.
FT   REPEAT      234    257       LRR 8.
FT   REPEAT      258    281       LRR 9.
FT   REPEAT      282    305       LRR 10.
FT   REPEAT      306    329       LRR 11.
FT   REPEAT      330    353       LRR 12.
FT   REPEAT      354    377       LRR 13.
FT   REPEAT      378    401       LRR 14.
FT   REPEAT      402    425       LRR 15.
FT   REPEAT      426    449       LRR 16.
FT   REPEAT      450    473       LRR 17.
FT   REPEAT      474    497       LRR 18.
FT   REPEAT      498    521       LRR 19.
FT   REPEAT      522    536       LRR 20.
FT   CARBOHYD     64     64       POTENTIAL.
FT   CARBOHYD     85     85       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    368    368       POTENTIAL.
FT   CARBOHYD    515    515       POTENTIAL.
FT   CARBOHYD    580    580       POTENTIAL..
SQ   SEQUENCE   605 AA;  66110 MW;  5DF04D42 CRC32;

  Query Match           18.6%;  Score 185;  DB 1;  Length 605;
  Best Local Similarity  34.8%;  Pred. No. 2.67e-17;
  Matches   31;  Conservative   28;  Mismatches  27;  Indels   3;  Gaps   3;

Db    187 DAAFRGLGGLRELVLAGNRLAYLQPALF-SGLAELRELDLSRNALRA-IKANVFAQLPRL 244
          ::|| | ::::||:|:||:|  ::   | :||: |: | | |: | : :  : || |: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    245 QKLYLDRNLIAAVAPGAFLGLKALRWLDL 273
          : | |  | |::::|||| | :|  ::|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT    5
ID   ALS_HUMAN      STANDARD;      PRT;   605 AA.
AC   P35858;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 92357025.
RA   LEONG S.R., BAXTER R.C., CAMERATO T., DAI J., WOOD W.I.;
RT   "Structure and functional expression of the acid-labile subunit of
RT   the insulin-like growth factor-binding protein complex.";
RL   MOL. ENDOCRINOL. 6:870-876(1992).
RN   [2]
RP   SEQUENCE OF 28-35.
RX   MEDLINE; 89308584.
RA   BAXTER R.C., MARTIN J.L., BENIAC V.A.;
RT   "High molecular weight insulin-like growth factor binding protein
RT   complex. Purification and properties of the acid-labile subunit from
RT   human serum.";
RL   J. BIOL. CHEM. 264:11843-11848(1989).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M86826; G184808; -.
DR   PIR; A41915; A41915.
DR   MIM; 601489; -.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL       1     27
FT   CHAIN       28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                               PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN       79    536       LEUCINE-RICH REPEATS.
FT   REPEAT       79     89       LRR 1.
FT   REPEAT       90    113       LRR 2.
FT   REPEAT      114    137       LRR 3.
FT   REPEAT      138    161       LRR 4.
FT   REPEAT      162    185       LRR 5.
FT   REPEAT      186    209       LRR 6.
FT   REPEAT      210    233       LRR 7.
FT   REPEAT      234    257       LRR 8.
FT   REPEAT      258    281       LRR 9.
FT   REPEAT      282    305       LRR 10.
FT   REPEAT      306    329       LRR 11.
FT   REPEAT      330    353       LRR 12.
FT   REPEAT      354    377       LRR 13.
FT   REPEAT      378    401       LRR 14.
FT   REPEAT      402    425       LRR 15.
FT   REPEAT      426    449       LRR 16.
FT   REPEAT      450    473       LRR 17.
FT   REPEAT      474    497       LRR 18.
FT   REPEAT      498    521       LRR 19.
FT   REPEAT      522    536       LRR 20.
FT   CARBOHYD     64     64       POTENTIAL.
FT   CARBOHYD     85     85       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    368    368       POTENTIAL.
FT   CARBOHYD    515    515       POTENTIAL.
FT   CARBOHYD    580    580       POTENTIAL.
SQ   SEQUENCE   605 AA;  66034 MW;  B5027E19 CRC32;

  Query Match          18.5%;  Score 184;  DB 1;  Length 605;
  Best Local Similarity  34.8%;  Pred. No. 4.20e-17;
  Matches   31;  Conservative   27;  Mismatches  28;  Indels   3;  Gaps   3;

Db    187 DAAFRGLGSLRELVLAGNRLAYLQPALF-SGLAELRELDLSRNALRA-IKANVFVQLPRL 244
          ::|| | :|::||:|:||:|  ::   | :||: |: | | |: | : :  : |  |: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    245 QKLYLDRNLIAAVAPGAFLGLKALRWLDL 273
          : | |  | |::::||||  | :|  ::|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT    6
ID   CBP8_HUMAN     STANDARD;      PRT;   536 AA.
AC   P22792;
DT   01-AUG-1991 (REL. 19, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CARBOXYPEPTIDASE N 83 KD CHAIN (CARBOXYPEPTIDASE N REGULATORY
DE   SUBUNIT) (FRAGMENT).
GN   CPN2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 90094386.
RA   TAN F., WEERASINGHE D.K., SKIDGEL R.A., TAMEI H., KAUL R.K.,
RA   RONINSON I.B., SCHILLING J.W., ERDOES E.G.;
RT   "The deduced protein sequence of the human carboxypeptidase N high
RT   molecular weight subunit reveals the presence of leucine-rich tandem
RT   repeats.";
RL   J. BIOL. CHEM. 265:13-19(1990).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE; 88309120.
RA   SKIDGEL R.A., BENNETT C.D., SCHILLING J.W., TAN F., WEERASINGHE D.K.,
RA   ERDOES E.G.;
RT   "Amino acid sequence of the N-terminus and selected tryptic peptides
RT   of the active subunit of human plasma carboxypeptidase N: comparison
RT   with other carboxypeptidases.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 154:1323-1329(1988).
CC   -!- FUNCTION: THE 83 KD SUBUNIT BINDS AND STABILIZES THE CATALYTIC
CC       SUBUNIT AT 37 DEGREES CELSIUS AND KEEPS IT IN CIRCULATION. UNDER
CC       SOME CIRCUMSTANCES IT MAY BE AN ALLOSTERIC MODIFIER OF THE
CC       CATALYTIC SUBUNIT.
CC   -!- SUBUNIT: TETRAMER OF TWO CATALYTIC CHAINS AND TWO GLYCOSYLATED
CC       INACTIVE CHAINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGION (POTENTIAL).
CC   -!- PTM: WHETHER OR NOT ANY CYS RESIDUES PARTICIPATE IN INTRACHAIN
CC       BONDS IS UNKNOWN, BUT THEY DO NOT FORM INTERCHAIN DISULFIDE BONDS
CC       WITH THE 50 KD CATALYTIC SUBUNIT.
CC   -!- DISEASE: A COMPLETE ABSENCE OF THE ENZYME IS NOT CONSIDERED TO BE
CC       COMPATIBLE WITH LIFE.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   -!- SIMILARITY: SOME, TO E.COLI YDDK.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J05158; G179936; -.
DR   PIR; A34901; A34901.
DR   MIM; 603104; -.
DR   PFAM; PF00560; LRR; 7.
KW   REPEAT; LEUCINE-REPEAT; GLYCOPROTEIN.
FT   NON_TER      1      1
FT   DOMAIN      68    355       LEUCINE-RICH REPEATS.
FT   REPEAT      68     91       LRR 1.
FT   REPEAT      92    115       LRR 2.
FT   REPEAT     116    139       LRR 3.
FT   REPEAT     140    163       LRR 4.
FT   REPEAT     164    187       LRR 5.
FT   REPEAT     188    211       LRR 6.
FT   REPEAT     212    235       LRR 7.
FT   REPEAT     236    259       LRR 8.
FT   REPEAT     260    283       LRR 9.
FT   REPEAT     284    307       LRR 10.
FT   REPEAT     308    331       LRR 11.
FT   REPEAT     332    355       LRR 12.
FT   DOMAIN     359    379       THR/SER-RICH.
FT   CARBOHYD    53     53       POTENTIAL.
FT   CARBOHYD    90     90       POTENTIAL.
FT   CARBOHYD    98     98       POTENTIAL.
FT   CARBOHYD   207    207       POTENTIAL.
FT   CARBOHYD   245    245       POTENTIAL.
FT   CARBOHYD   327    327       POTENTIAL.
FT   CARBOHYD   338    338       POTENTIAL.
FT   CARBOHYD   495    495       POTENTIAL.
SQ   SEQUENCE   536 AA;  58649 MW;  C4413E03 CRC32;
```

```
  Query Match          18.4%;  Score 183;  DB 1;  Length 536;
  Best Local Similarity  37.4%;  Pred. No. 6.59e-17;
  Matches   34;  Conservative   22;  Mismatches  34;  Indels   1;  Gaps   1;

Db    117 EGLFQHLAALESLHLQGNQLQALPRRLFQP-LTHLKTLNLAQNLLAQLPEELFHPLTSLQ 175
          || |   |::: | | ||||::: | |:  |: |||| |  ||:: :::: |  |:|::
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    176 TLKLSNNALSGLPQGVFGKLGSLQELFLDSN 206
          | | :| :: :  | |  | ||  : | ||
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSN 91


RESULT   7
ID   ALS_RAT       STANDARD;      PRT;   603 AA.
AC   P35859;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93038676.
RA   DAI J., BAXTER R.C.;
RT   "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT   growth factor binding protein complex.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 188:304-309(1992).
RN   [2]
RP   SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC   STRAIN=WISTAR; TISSUE=SERUM;
RX   MEDLINE; 94130835.
RA   BAXTER R.C., DAI J.;
RT   "Purification and characterization of the acid-labile subunit of rat
RT   serum insulin-like growth factor binding protein complex.";
RL   ENDOCRINOLOGY 134:848-852(1994).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC       AND LIVER.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S46785; E64972; -.
DR   PIR; JC1282; JC1282.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL        1     23
FT   CHAIN        24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN       79    535       LEUCINE-RICH REPEATS.
FT   REPEAT       79     89       LRR 1.
FT   REPEAT       90    113       LRR 2.
FT   REPEAT      114    137       LRR 3.
FT   REPEAT      138    161       LRR 4.
FT   REPEAT      162    185       LRR 5.
FT   REPEAT      186    209       LRR 6.
FT   REPEAT      210    233       LRR 7.
FT   REPEAT      234    257       LRR 8.
FT   REPEAT      258    281       LRR 9.
FT   REPEAT      282    305       LRR 10.
FT   REPEAT      306    329       LRR 11.
FT   REPEAT      330    353       LRR 12.
FT   REPEAT      354    377       LRR 13.
FT   REPEAT      378    401       LRR 14.
FT   REPEAT      402    425       LRR 15.
FT   REPEAT      426    449       LRR 16.
FT   REPEAT      450    473       LRR 17.
FT   REPEAT      474    497       LRR 18.
FT   REPEAT      498    521       LRR 19.
FT   REPEAT      522    535       LRR 20.
FT   CARBOHYD     64     64       POTENTIAL.
FT   CARBOHYD     85     85       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    368    368       POTENTIAL.
FT   CARBOHYD    515    515       POTENTIAL.
FT   CARBOHYD    578    578       POTENTIAL.
FT   CARBOHYD    586    586       POTENTIAL.
SQ   SEQUENCE   603 AA;  66811 MW;  5BB22D53 CRC32;

  Query Match          18.0%;  Score 179;  DB 1;  Length 603;
  Best Local Similarity  37.5%;  Pred. No. 3.97e-16;
  Matches   39;  Conservative   22;  Mismatches  41;  Indels   2;  Gaps   2;

Db    356 GAFSGLFNVAVMNLSGNCLRSLPERVFQG-LDKLHSLHLEHSCLGHVRLHTFAGLSGLRR 414
          |||:|  :|  : |:|| | ::  | |:| |  | :| |  : :| |   ||||||::|
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    415 LFLRDNSISSIEEQSLAGLSELLELDLTTNRLTHLPRQLFQGLG 458
          | | || |::|   ::: |  |  ::| :| :     :|  |||
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINLLSNPFNC-NCHLGAGLG 104


RESULT   8
ID   GPV_RAT       STANDARD;      PRT;   567 AA.
AC   O08770;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RX   MEDLINE; 97275136.
RA   RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,
RA   GRUNERT P., LOEW D., VAN DORSSELAER A., CAZENAVE J.-P., LANZA F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   BLOOD 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z69594; E222201; -.
DR   PFAM; PF00560; LRR; 8.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17   522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   523   543       POTENTIAL.
FT   DOMAIN     544   567       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55   415       LEUCINE-RICH REPEATS.
FT   REPEAT      55    78       LRR 1.
FT   REPEAT      79   102       LRR 2.
FT   REPEAT     103   126       LRR 3.
FT   REPEAT     127   150       LRR 4.
FT   REPEAT     151   174       LRR 5.
FT   REPEAT     175   198       LRR 6.
FT   REPEAT     199   222       LRR 7.
FT   REPEAT     223   246       LRR 8.
FT   REPEAT     247   270       LRR 9.
FT   REPEAT     271   294       LRR 10.
FT   REPEAT     295   318       LRR 11.
FT   REPEAT     319   343       LRR 12.
FT   REPEAT     346   367       LRR 13.
FT   REPEAT     368   391       LRR 14.
FT   REPEAT     392   415       LRR 15.
FT   CARBOHYD    51    51       POTENTIAL.
FT   CARBOHYD   181   181       POTENTIAL.
FT   CARBOHYD   243   243       POTENTIAL.
FT   CARBOHYD   298   298       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   CARBOHYD   385   385       POTENTIAL.
FT   CARBOHYD   498   498       POTENTIAL.
SQ   SEQUENCE   567 AA;  63344 MW;  ABAEC91D CRC32;

  Query Match          17.7%;  Score 176;  DB 1;  Length 567;
  Best Local Similarity  31.2%;  Pred. No. 1.51e-15;
  Matches   34;  Conservative   26;  Mismatches  47;  Indels   2;  Gaps   2;

Db    333 GMFHGLTELRVLAVHTNALEELPEDALRG-LGRLRQVSLRHNRLRALPRTLFRNLSSLVT 391
          | |:| : :: | :  | || :   ::|| |: |: : || | :  ::   |  |||:
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    392 VQLEHNQLKTLPGDVFAALPQLTRVLLGHNPWLCDCGLWPFL-QWLRHH 439
          : |  |:: |:    |::|  |: : |  ||  |:| | : |  ||| :
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKR 110


RESULT   9
ID   CONN_DROME     STANDARD;      PRT;   682 AA.
AC   Q01819;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CONNECTIN PRECURSOR.
GN   CON.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92370678.
RA   NOSE A., MAHAJAN V.B., GOODMAN C.S.;
RT   "Connectin: a homophilic cell adhesion molecule expressed on a subset
RT   of muscles and the motoneurons that innervate them in Drosophila.";
RL   CELL 70:553-567(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 93202002.
RA   GOULD A.P., WHITE R.A.H.;
RT   "Connectin, a target of homeotic gene control in Drosophila.";
RL   DEVELOPMENT 116:1163-1174(1992).
CC   -!- FUNCTION: CELL ADHESION PROTEIN INVOLVED IN TARGET RECOGNITION
CC       DURING NEUROMUSCULAR DEVELOPMENT. MEDIATES HOMOPHILIC CELLULAR
CC       ADHESION.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN ABDOMINAL AND
CC       THORACIC SEGMENT MUSCLE AND MOTORNEURON CELLS.
CC   -!- DEVELOPMENTAL STAGE: EMBRYO.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN:10.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M96647; G157084; -.
DR   EMBL; X68701; G7738; -.
DR   PIR; S28464; S28464.
DR   PIR; A43318; A43318.
DR   FLYBASE; FBgn0005775; Con.
DR   PFAM; PF00560; LRR; 5.
KW   CELL ADHESION; DEVELOPMENTAL PROTEIN; EMBRYO; SIGNAL; GPI-ANCHOR;
KW   LEUCINE-REPEAT; REPEAT.
FT   SIGNAL       1    24
FT   CHAIN       25   665       CONNECTIN.
FT   PROPEP     666   682       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN     142   381       LEUCINE-RICH REPEATS.
FT   REPEAT     142   165       LRR 1.
FT   REPEAT     166   189       LRR 2.
FT   REPEAT     190   213       LRR 3.
FT   REPEAT     214   237       LRR 4.
FT   REPEAT     238   261       LRR 5.
FT   REPEAT     262   285       LRR 6.
FT   REPEAT     286   299       LRR 7.
FT   REPEAT     300   322       LRR 8.
FT   REPEAT     324   347       LRR 9.
FT   REPEAT     348   381       LRR 10.
FT   LIPID      665   665       GPI-ANCHOR (POTENTIAL).
FT   CONFLICT   631   631       E -> G (IN REF. 2).
FT   CONFLICT   674   677       QVAL -> VALM (IN REF. 2).
SQ   SEQUENCE   682 AA;  75992 MW;  3E15592A CRC32;

  Query Match          17.3%;  Score 172;  DB 1;  Length 682;
  Best Local Similarity  39.1%;  Pred. No. 8.93e-15;
  Matches   36;  Conservative   19;  Mismatches  34;  Indels   3;  Gaps   3;

Db    263 EGLFADMARLTFLNLAHNQINVLTSEIFRG-LGNLNVLKLTRNNL-NFIGDTVFAELWSL 320
          || |   | :  | |: ||:: : :  ||| |: |: |:| |:||   :::  || | |:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSV 59

Db    321 SELELDDNRIERISERALDGLNTLKTLNLRNN 352
            | | ||||  |:  |:  | :| |:|| :|
Qy     60 RLLSLYDNRITTITPGAFTTLVSLSTINLLSN 91


RESULT   10
ID   GPV_MOUSE      STANDARD;      PRT;   567 AA.
AC   O08742;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=C57BL/6; TISSUE=LIVER;
RX   MEDLINE; 97275136.
RA   RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,
RA   GRUNERT P., LOEW D., VAN DORSSELAER A., CAZENAVE J.-P., LANZA F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   BLOOD 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z69595; E222202; -.
DR   MGD; MGI:1096363; GP5.
DR   PFAM; PF00560; LRR; 7.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17   522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   523   543       POTENTIAL.
FT   DOMAIN     544   567       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55   415       LEUCINE-RICH REPEATS.
FT   REPEAT      55    78       LRR 1.
FT   REPEAT      79   102       LRR 2.
FT   REPEAT     103   126       LRR 3.
FT   REPEAT     127   150       LRR 4.
FT   REPEAT     151   174       LRR 5.
FT   REPEAT     175   198       LRR 6.
FT   REPEAT     199   222       LRR 7.
FT   REPEAT     223   246       LRR 8.
FT   REPEAT     247   270       LRR 9.
FT   REPEAT     271   294       LRR 10.
FT   REPEAT     295   318       LRR 11.
FT   REPEAT     319   343       LRR 12.
FT   REPEAT     346   367       LRR 13.
FT   REPEAT     368   391       LRR 14.
FT   REPEAT     392   415       LRR 15.
FT   CARBOHYD    51    51       POTENTIAL.
FT   CARBOHYD    67    67       POTENTIAL.
FT   CARBOHYD   181   181       POTENTIAL.
FT   CARBOHYD   243   243       POTENTIAL.
FT   CARBOHYD   298   298       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   CARBOHYD   385   385       POTENTIAL.
SQ   SEQUENCE   567 AA;  63467 MW;  3AE7515E CRC32;

  Query Match          17.2%;  Score 171;  DB 1;  Length 567;
  Best Local Similarity  33.7%;  Pred. No. 1.39e-14;
  Matches   29;  Conservative   21;  Mismatches  34;  Indels   2;  Gaps   2;

Db    286 FGEMAGLRELWLNGTHLSTLPAAAFRN-LSGLQTLGLTRNPRLSALPRGVFQGLRELRVL 344
          |   |:::|| | | :| |: : :||  |||| || | |:  :: ::   | ||  :|:|
Qy      4 FNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLML-RSNLIGCVSNDTFAGLSSVRLL 62

Db    345 GLHTNALAELRDDALRGLGHLRQVSL 370
          :|  | :: :   |:  |  |  ::|
Qy     63 SLYDNRITTITPGAFTTLVSLSTINL 88
```

```
RESULT   11
ID   PKD1_HUMAN     STANDARD;      PRT;  4303 AA.
AC   P98161;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   POLYCYSTIN PRECURSOR (AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE
DE   PROTEIN 1).
GN   PKD1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95254638.
RA   GLUECKSMANN-KUIS M.A., TAYBER O., WOOLF E.A., BOUGUELERET L.,
RA   DENG N., ALPERIN G.D., IRIS F., HAWKINS F., MUNRO C., LAKEY N.,
RA   DUYK G., SCHNEIDER M.C., GENG L., ZHANG F., ZHAO Z., TOROSIAN S.,
RA   REEDERS S.T., BORK P., POHLSCHMIDT M., LOEHNING C., KRAUS B.,
RA   NOWICKA U., LEUNG A.L.S., FRISCHAUF A.-M.;
RT   "Polycystic kidney disease: the complete structure of the PKD1 gene
RT   and its protein. The International Polycystic Kidney Disease
RT   Consortium.";
RL   CELL 81:289-298(1995).
RN   [2]
RP   SEQUENCE OF 2769-4303 FROM N.A.
RX   MEDLINE; 94273192.
RA   WARD C.J., PERAL B., HUGHES J., THOMAS S., GAMBLE V.,
RA   MACCARTHY A.B., SLOANE-STANLEY J., BUCKLE V.J., KEARNEY L.,
RA   HIGGS D.R., RATCLIFFE P.J., HARRIS P.C., ROELFSEMA J.H.,
RA   SPRUIT L.L., SARIS J.J., DAUWERSE H.G., PETERS D.J.M.,
RA   BREUNING M.H., NELLIST M., BROOK-CARTER P.T., MAHESHWAR M.M.,
RA   CORDEIRO I., SANTOS H., CABRAL P., SAMPSON J.R., JANSSEN B.,
RA   HESSELING-JANSSEN A.L.W., VAN DEN OUWELAND A.M.W., EUSSEN B.,
RA   VERHOEF S., LINDHOUT D., HALLEY D.J.J.;
RT   "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT   lies within a duplicated region on chromosome 16. The European
RT   Polycystic Kidney Disease Consortium.";
RL   CELL 77:881-894(1994).
RN   [3]
RP   VARIANT ADPKD 3748-ARG-VAL-3752 DEL, AND VARIANT ASP-3632.
RX   MEDLINE; 96108969.
RA   PERAL B., SAN MILLAN J.L., ONG A.C.M., GAMBLE V., WARD C.J.,
RA   STRONG C., HARRIS P.C.;
RT   "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
RT   gene reveals six novel mutations.";
RL   AM. J. HUM. GENET. 58:86-96(1996).
RN   [4]
RP   VARIANT ALA-4058.
RX   MEDLINE; 97295081.
RA   CONSTANTINIDES R., XENOPHONTOS S., NEOPHYTOU P., NOMURA S.,
RA   PIERIDES A., CONSTANTINOU-DELTAS C.D.;
RT   "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
RT   polycystic kidney disease 1 gene: evolution of alleles.";
RL   HUM. GENET. 99:644-647(1997).
RN   [5]
RP   VARIANTS T-2760;P-2761;V-2763;T-2764;Q-2791;T-2826;L-3008 & L-3064.
RX   MEDLINE; 97449169.
RA   WATNICK T.J., PIONTEK K.B., CORDAL T.M., WEBER H., GANDOLPH M.A.,
RA   QIAN F., LENS X.M., NEUMANN H.P.H., GERMINO G.G.;
RT   "An unusual pattern of mutation in the duplicated portion of PKD1 is
RT   revealed by use of a novel strategy for mutation detection.";
RL   HUM. MOL. GENET. 6:1473-1481(1997).
RN   [6]
RP   VARIANT ADPKD THR-3678.
RX   MEDLINE; 97403939.
RA   TURCO A.E., ROSSETTI S., BRESIN E., ENGLISCH S., CORRA S.,
RA   PIGNATTI P.F.;
RT   "Three novel mutations of the PKD1 gene in Italian families with
RT   autosomal dominant polycystic kidney disease.";
RL   HUM. MUTAT. 10:164-167(1997).
RN   [7]
RP   VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
```

```
RX   MEDLINE; 98180892.
RA   DANIELLS C., MAHESHWAR M., LAZAROU L., DAVIES F., COLES G., RAVINE D.;
RT   "Novel and recurrent mutations in the PKD1 (polycystic kidney
RT   disease) gene.";
RL   HUM. GENET. 102:216-220(1998).
CC   -!- FUNCTION: INVOLVED IN ADHESIVE PROTEIN-PROTEIN AND PROTEIN-
CC       CARBOHYDRATE INTERACTIONS.
CC   -!- DISEASE: DEFECTS IN PKD1 ARE THE CAUSE OF AUTOSOMAL DOMINANT
CC       POLYCYSTIC KIDNEY DISEASE (ADPKD), A COMMON AUTOSOMAL DOMINANT
CC       GENETIC DISEASE AFFECTING ABOUT 1 OUT 1000 INDIVIDUALS. IT IS
CC       CHARACTERIZED BY PROGRESSIVE FORMATION AND ENLARGEMENT OF CYSTS IN
CC       BOTH KIDNEYS, TYPICALLY LEADING TO END-STAGE RENAL DISEASE IN
CC       ADULT LIFE. CYSTS ALSO OCCURS IN THE LIVER AND OTHER ORGANS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 2, FRAMED BY A
CC       LRR N-FLANK AND LRR C-FLANK.
CC   -!- SIMILARITY: CONTAINS 14 POLYCYSTIC KIDNEY DISEASE DOMAINS (PKD).
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN (ATYPICAL,
CC       THE POTENTIAL CALCIUM-BINDING SITE IS MISSING).
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U24497; G799335; -.
DR   MIM; 601313; -.
DR   MIM; 173900; -.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PFAM; PF00560; LRR; 2.
DR   PFAM; PF00801; PKD; 16.
KW   SIGNAL; LEUCINE-REPEAT; LECTIN; REPEAT; GLYCOPROTEIN; TRANSMEMBRANE;
KW   DISEASE MUTATION; POLYMORPHISM.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24   4303       POLYCYSTIN.
FT   DOMAIN     638    671       LDL-RECEPTOR CLASS A.
FT   DOMAIN      32     65       LRR N-FLANK.
FT   DOMAIN      72    119       LEUCINE-RICH REPEATS.
FT   REPEAT      72     94       LRR 1.
FT   REPEAT      97    119       LRR 2.
FT   DOMAIN     123    177       LRR C-FLANK.
FT   REPEAT     282    353       PKD 1.
FT   DOMAIN     405    534       C-TYPE LECTIN.
FT   REPEAT     844    929       PKD 2.
FT   REPEAT     930   1014       PKD 3.
FT   DOMAIN    1032   2142       13 X 80 AA REPEATS.
FT   REPEAT    1032   1124       PKD 4.
FT   REPEAT    1138   1209       PKD 5.
FT   REPEAT    1221   1292       PKD 6.
FT   REPEAT    1305   1377       PKD 7.
FT   REPEAT    1390   1463       PKD 8.
FT   REPEAT    1477   1545       PKD 9.
FT   REPEAT    1559   1629       PKD 10.
FT   REPEAT    1643   1715       PKD 11.
FT   REPEAT    1729   1799       PKD 12.
FT   REPEAT    1815   1883       PKD 13.
FT   REPEAT    1898   1968       PKD 14.
FT   REPEAT    1983   2058       PKD 15.
FT   REPEAT    2071   2142       PKD 16.
FT   TRANSMEM  2216   2236       POTENTIAL.
FT   TRANSMEM  2546   2566       POTENTIAL.
FT   TRANSMEM  2580   2600       POTENTIAL.
FT   TRANSMEM  2693   2713       POTENTIAL.
FT   TRANSMEM  2808   2828       POTENTIAL.
FT   TRANSMEM  3075   3095       POTENTIAL.
FT   TRANSMEM  3283   3303       POTENTIAL.
FT   TRANSMEM  3324   3344       POTENTIAL.
FT   TRANSMEM  3560   3580       POTENTIAL.
FT   TRANSMEM  3583   3603       POTENTIAL.
FT   TRANSMEM  3674   3694       POTENTIAL.
FT   TRANSMEM  3897   3917       POTENTIAL.
FT   TRANSMEM  3939   3959       POTENTIAL.
FT   TRANSMEM  3980   4000       POTENTIAL.
FT   TRANSMEM  4028   4048       POTENTIAL.
FT   TRANSMEM  4055   4075       POTENTIAL.
FT   TRANSMEM  4086   4106       POTENTIAL.
FT   DISULFID   640    653       BY SIMILARITY.
FT   DISULFID   647    665       BY SIMILARITY.
FT   DISULFID   660    669       BY SIMILARITY.
FT   CARBOHYD    50     50       POTENTIAL.
FT   CARBOHYD    89     89       POTENTIAL.
FT   CARBOHYD   116    116       POTENTIAL.
FT   CARBOHYD   121    121       POTENTIAL.
FT   CARBOHYD   187    187       POTENTIAL.
FT   CARBOHYD   621    621       POTENTIAL.
FT   CARBOHYD   632    632       POTENTIAL.
FT   CARBOHYD   746    746       POTENTIAL.
FT   CARBOHYD   810    810       POTENTIAL.
FT   CARBOHYD   841    841       POTENTIAL.
FT   CARBOHYD   854    854       POTENTIAL.
FT   CARBOHYD   890    890       POTENTIAL.
FT   CARBOHYD   921    921       POTENTIAL.
FT   CARBOHYD  1004   1004       POTENTIAL.
FT   CARBOHYD  1010   1010       POTENTIAL.
FT   CARBOHYD  1034   1034       POTENTIAL.
FT   CARBOHYD  1072   1072       POTENTIAL.
FT   CARBOHYD  1090   1090       POTENTIAL.
FT   CARBOHYD  1113   1113       POTENTIAL.
FT   CARBOHYD  1178   1178       POTENTIAL.
FT   CARBOHYD  1194   1194       POTENTIAL.
FT   CARBOHYD  1240   1240       POTENTIAL.
FT   CARBOHYD  1269   1269       POTENTIAL.
FT   CARBOHYD  1336   1336       POTENTIAL.
FT   CARBOHYD  1348   1348       POTENTIAL.
FT   CARBOHYD  1382   1382       POTENTIAL.
FT   CARBOHYD  1450   1450       POTENTIAL.
FT   CARBOHYD  1455   1455       POTENTIAL.
FT   CARBOHYD  1474   1474       POTENTIAL.
FT   CARBOHYD  1518   1518       POTENTIAL.
FT   CARBOHYD  1541   1541       POTENTIAL.
FT   CARBOHYD  1554   1554       POTENTIAL.
FT   CARBOHYD  1563   1563       POTENTIAL.
FT   CARBOHYD  1647   1647       POTENTIAL.
FT   CARBOHYD  1661   1661       POTENTIAL.
FT   CARBOHYD  1733   1733       POTENTIAL.
FT   CARBOHYD  1791   1791       POTENTIAL.
FT   CARBOHYD  1834   1834       POTENTIAL.
FT   CARBOHYD  1867   1867       POTENTIAL.
FT   CARBOHYD  1880   1880       POTENTIAL.
FT   CARBOHYD  1991   1991       POTENTIAL.
FT   CARBOHYD  2050   2050       POTENTIAL.
FT   CARBOHYD  2074   2074       POTENTIAL.
FT   CARBOHYD  2125   2125       POTENTIAL.
FT   CARBOHYD  2248   2248       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          16.8%;  Score 167;  DB 1;  Length 4303;
  Best Local Similarity  32.2%;  Pred. No. 8.06e-14;
  Matches   39;  Conservative   27;  Mismatches  49;  Indels   6;  Gaps   6;

Db     52 SGRGLRTL-GPALRIP-ADATELDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEG 109
          :| | |: | ::|   :    | :  ||: :    :| ||:: | : :|:|:|:  |
Qy     16 TGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPG 75

Db    110 IFANLFNLSEINLSGNPFECDCGL-AWLPQWAEEQQVRVVQPEAATCAGPGSLAGQPLLG 168
           |: | :|| ||| :|||:|:| | | |  |    : |:| :  : |  |  |   |: |
Qy     76 AFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLR-KR-RIV-SGNPRCQKPFFLKEIPIQG 132

Db    169 I 169
          :
Qy    133 V 133
```

```
RESULT  12
ID   ALS_MOUSE      STANDARD;      PRT;   603 AA.
AC   P70389;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS OR ALBS.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE; 96413591.
RA   BOISCLAIR Y.R., SETO D., HSIEH S., HURST K.R., OOI G.T.;
RT   "Organization and chromosomal localization of the gene encoding the
RT   mouse acid labile subunit of the insulin-like growth factor binding
RT   complex.";
RP   PROC. NATL. ACAD. SCI. U.S.A. 93:10028-10033(1996).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66900; G1621613; -.
DR   MGD; MGI:107973; IGFALS.
DR   PFAM; PF00560; LRR; 10.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL       1    23       BY SIMILARITY.
FT   CHAIN       24   603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                              PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN      79   535       LEUCINE-RICH REPEATS.
FT   REPEAT      79    89       LRR 1.
FT   REPEAT      90   113       LRR 2.
FT   REPEAT     114   137       LRR 3.
FT   REPEAT     138   161       LRR 4.
FT   REPEAT     162   185       LRR 5.
FT   REPEAT     186   209       LRR 6.
FT   REPEAT     210   233       LRR 7.
FT   REPEAT     234   257       LRR 8.
FT   REPEAT     258   281       LRR 9.
FT   REPEAT     282   305       LRR 10.
FT   REPEAT     306   329       LRR 11.
FT   REPEAT     330   353       LRR 12.
FT   REPEAT     354   377       LRR 13.
FT   REPEAT     378   401       LRR 14.
FT   REPEAT     402   425       LRR 15.
FT   REPEAT     426   449       LRR 16.
FT   REPEAT     450   473       LRR 17.
FT   REPEAT     474   497       LRR 18.
FT   REPEAT     498   521       LRR 19.
FT   REPEAT     522   535       LRR 20.
FT   CARBOHYD    64    64       POTENTIAL.
FT   CARBOHYD    85    85       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   368   368       POTENTIAL.
FT   CARBOHYD   515   515       POTENTIAL.
FT   CARBOHYD   578   578       POTENTIAL.
FT   CARBOHYD   586   586       POTENTIAL.
SQ   SEQUENCE   603 AA;  66959 MW;  11ADB606 CRC32;

  Query Match          16.7%;  Score 166;  DB 1;  Length 603;
  Best Local Similarity  32.1%;  Pred. No. 1.25e-13;
  Matches   34;  Conservative   27;  Mismatches  43;  Indels   2;  Gaps   2;

Db    283 EDTFPGLLGLHVLRLAHNAITSLRPRTFKD-LHFLEELQLGHNRIRQLGEKTFEGLGQLE 341
          | :| |  ::: | |: | : ::: | |:  |  |  | |  | |  ::: || ||: :
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    342 VLTLNDNQIHEVKVGAFFGLFNVAVMNLSGNCLRSLPEHVFQGLGR 387
          :|:| ||:|  :  |||  | ::: :|| :| :     |:  |||:
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNC-HLGAGLGK 105


RESULT  13
ID   GPCR_LYMST     STANDARD;      PRT;  1115 AA.
AC   P46023;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   G-PROTEIN COUPLED RECEPTOR GRL101 PRECURSOR.
OS   LYMNAEA STAGNALIS (GREAT POND SNAIL).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PULMONATA; BASOMMATOPHORA;
OC   LYMNAEIDAE; LYMNAEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CNS;
RX   MEDLINE; 94255418.
RA   TENSEN C.P., VAN KESTEREN E.R., PLANTA R.J., COX K.J.A., BURKE J.F.,
RA   VAN HEERIKHUIZEN H., VREUGDENHIL E.;
RT   "A G protein-coupled receptor with low density lipoprotein-binding
RT   motifs suggests a role for lipoproteins in G-linked signal
RT   transduction.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 91:4816-4820(1994).
CC   -!- FUNCTION: MIGHT DIRECTLY TRANSDUCE SIGNALS CARRIED BY LARGE
CC       EXTRACELLULAR (LIPO)PROTEIN(COMPLEXE)S INTO NEURONAL EVENTS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN A SMALL NUMBER OF
CC       NEURONS WITHIN THE CENTRAL NERVOUS SYSTEM AND TO A LESSER EXTENT
CC       IN THE HEART.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 12 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 6.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z23104; G438129; -.
DR   PIR; S40241; S40241.
DR   GCRDB; GCR_0860; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR   PROSITE; PS01209; LDLRA_1; 6.
DR   PROSITE; PS50068; LDLRA_2; 11.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00057; ldl_recept_a; 11.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P01130; 1AJJ.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
KW   SIGNAL.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25  1115       G-PROTEIN COUPLED RECEPTOR GRL101.
FT   DOMAIN      25   767       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   768   788       1 (POTENTIAL).
FT   DOMAIN     789   801       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   802   822       2 (POTENTIAL).
```

```
FT   DOMAIN      823    857       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    858    878       3 (POTENTIAL).
FT   DOMAIN      879    887       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    888    908       4 (POTENTIAL).
FT   DOMAIN      909    941       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    942    962       5 (POTENTIAL).
FT   DOMAIN      963    988       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    989   1009       6 (POTENTIAL).
FT   DOMAIN     1010   1017       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1018   1038       7 (POTENTIAL).
FT   DOMAIN     1039   1115       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       32    523       12 X 40 AA APPROXIMATE TANDEM REPEATS
FT                                SIMILAR TO THE LDL-RECEPTOR CLASS A.
FT   DOMAIN       36     79       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN       77    115       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      116    155       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      156    196       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      195    232       LDL-RECEPTOR CLASS A 5.
FT   DOMAIN      231    269       LDL-RECEPTOR CLASS A 6.
FT   DOMAIN      272    318       LDL-RECEPTOR CLASS A 7.
FT   DOMAIN      320    363       LDL-RECEPTOR CLASS A 8.
FT   DOMAIN      365    403       LDL-RECEPTOR CLASS A 9.
FT   DOMAIN      404    442       LDL-RECEPTOR CLASS A 10.
FT   DOMAIN      444    485       LDL-RECEPTOR CLASS A 11.
FT   DOMAIN      486    525       LDL-RECEPTOR CLASS A 12.
FT   DOMAIN      588    731       LEUCINE-RICH REPEATS.
FT   REPEAT      588    611       LRR 1.
FT   REPEAT      612    635       LRR 2.
FT   REPEAT      636    659       LRR 3.
FT   REPEAT      660    683       LRR 4.
FT   REPEAT      684    707       LRR 5.
FT   REPEAT      708    731       LRR 6.
FT   DISULFID     38     53       BY SIMILARITY.
FT   DISULFID     46     66       BY SIMILARITY.
FT   DISULFID     60     77       BY SIMILARITY.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     86    104       BY SIMILARITY.
FT   DISULFID     98    113       BY SIMILARITY.
FT   DISULFID    118    131       BY SIMILARITY.
FT   DISULFID    138    153       BY SIMILARITY.
FT   DISULFID    158    170       BY SIMILARITY.
FT   DISULFID    165    183       BY SIMILARITY.
FT   DISULFID    177    194       BY SIMILARITY.
FT   DISULFID    202    220       BY SIMILARITY.
FT   DISULFID    214    230       BY SIMILARITY.
FT   DISULFID    233    245       BY SIMILARITY.
FT   DISULFID    240    258       BY SIMILARITY.
FT   DISULFID    252    267       BY SIMILARITY.
FT   DISULFID    274    291       BY SIMILARITY.
FT   DISULFID    282    304       BY SIMILARITY.
FT   DISULFID    298    316       BY SIMILARITY.
FT   DISULFID    322    339       BY SIMILARITY.
FT   DISULFID    334    352       BY SIMILARITY.
FT   DISULFID    346    361       BY SIMILARITY.
FT   DISULFID    367    379       BY SIMILARITY.
FT   DISULFID    374    392       BY SIMILARITY.
FT   DISULFID    386    401       BY SIMILARITY.
FT   DISULFID    406    418       BY SIMILARITY.
FT   DISULFID    413    431       BY SIMILARITY.
FT   DISULFID    425    440       BY SIMILARITY.
FT   DISULFID    446    458       BY SIMILARITY.
FT   DISULFID    453    474       BY SIMILARITY.
FT   DISULFID    465    483       BY SIMILARITY.
FT   DISULFID    488    500       BY SIMILARITY.
FT   DISULFID    495    513       BY SIMILARITY.
FT   DISULFID    507    523       BY SIMILARITY.
FT   CARBOHYD     87     87       POTENTIAL.
FT   CARBOHYD    166    166       POTENTIAL.
FT   CARBOHYD    269    269       POTENTIAL.
FT   CARBOHYD    318    318       POTENTIAL.
FT   CARBOHYD    482    482       POTENTIAL.
FT   CARBOHYD    502    502       POTENTIAL.
FT   CARBOHYD    571    571       POTENTIAL.
FT   CARBOHYD    618    618       POTENTIAL.
FT   CARBOHYD    624    624       POTENTIAL.
FT   CARBOHYD    685    685       POTENTIAL.
SQ   SEQUENCE   1115 AA;  125865 MW;  2AEC245A CRC32;

  Query Match          15.9%;  Score 158;  DB 1;  Length 1115;
  Best Local Similarity  29.5%;  Pred. No. 4.01e-12;
  Matches   28;  Conservative   16;  Mismatches  50;  Indels   1;  Gaps   1;

Db    648 EDTFSSMIHLTVLDLSNQRLTHVYKNMFKG-LKQITVLNISRNQINSIDNGAFNNLANVR 706
          | :|::   :  | |: :|  |    |:| |  :  | :  | |  : | :|  |::||
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    707 LIDLSGNVIKDIGQKVFMGLPRLVELKTDSYRFCC 741
          |: |  | | |    | |  |  ::  |  | |
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNC 95


RESULT  14
ID   CHAO_DROME     STANDARD;      PRT;  1134 AA.
AC   P12024;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CHAOPTIN PRECURSOR (PHOTORECEPTOR CELL-SPECIFIC MEMBRANE PROTEIN).
GN   CHP OR CHT.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88135762.
RA   REINKE R., KRANTZ D.E., YEN D., ZIPURSKY S.L.;
RT   "Chaoptin, a cell surface glycoprotein required for Drosophila
RT   photoreceptor cell morphogenesis, contains a repeat motif found in
RT   yeast and human.";
RL   CELL 52:291-301(1988).
CC   -!- FUNCTION: REQUIRED FOR DROSOPHILA PHOTORECEPTOR CELL
CC       MORPHOGENESIS. MEDIATES HOMOPHILIC CELLULAR ADHESION.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR SURFACE OF R-CELL PLASMA
CC       MEMBRANE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED 24 HOURS AFTER INITIATION OF
CC       PHOTORECEPTOR CELL DIFFERENTIATION, PERSISTS THROUGH ADULTHOOD.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 41.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19017; G157098; -.
DR   EMBL; M19008; G157098; JOINED.
DR   EMBL; M19009; G157098; JOINED.
DR   EMBL; M19010; G157098; JOINED.
DR   EMBL; M19011; G157098; JOINED.
DR   EMBL; M19012; G157098; JOINED.
DR   EMBL; M19013; G157098; JOINED.
DR   EMBL; M19014; G157098; JOINED.
DR   EMBL; M19016; G157098; JOINED.
DR   PIR; A29944; A29944.
DR   FLYBASE; FBgn0000313; chp.
DR   PFAM; PF00560; LRR; 17.
KW   GLYCOPROTEIN; MEMBRANE; SIGNAL; REPEAT; LEUCINE-REPEAT; VISION.
FT   SIGNAL        1     29
FT   CHAIN        30   1134       CHAOPTIN.
FT   CARBOHYD     77     77       POTENTIAL.
FT   CARBOHYD    267    267       POTENTIAL.
FT   CARBOHYD    305    305       POTENTIAL.
```

```
FT   CARBOHYD    339    339       POTENTIAL.
FT   CARBOHYD    361    361       POTENTIAL.
FT   CARBOHYD    422    422       POTENTIAL.
FT   CARBOHYD    680    680       POTENTIAL.
FT   CARBOHYD    692    692       POTENTIAL.
FT   CARBOHYD    718    718       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    936    936       POTENTIAL.
FT   CARBOHYD    970    970       POTENTIAL.
FT   CARBOHYD   1012   1012       POTENTIAL.
FT   CARBOHYD   1104   1104       POTENTIAL.
SQ   SEQUENCE   1134 AA;  130719 MW;  B67A6363 CRC32;

  Query Match           15.0%;  Score 149;  DB 1;  Length 1134;
  Best Local Similarity  29.1%;  Pred. No. 1.85e-10;
  Matches   37;  Conservative   36;  Mismatches  47;  Indels   7;  Gaps   7;

Db    119 DDAFTGLERSLWELILPQNDLVEIPSKSLRH-LQKLRHLDLGYNHITHIQHDSFRGLEDS 177
          : || |   |: ||:|  |:|  : ::::|  |  |: | |  | |  : :|:| ||  |
Qy      1 EGAFNGAA-SVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGL-SS 58

Db    178 LQTLILRENCISQLMSHSFSGLLILETLDLSGNNLFEIDPNVFVDGMPR-L-TRLLLTDN 235
          :: | | :| |: : : :|: |: | |::| :|  |: : ::   |: : |  | ::: |
Qy     59 VRLLSLYDNRITTITPGAFTTLVSLSTINLLSNP-FNCNCHL-GAGLGKWLRKRRIVSGN 116

Db    236 ILSEIPY 242
            : |:
Qy    117 PRCQKPF 123

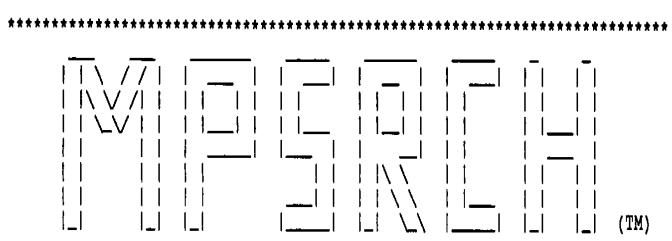

RESULT   15
ID   A2GL_HUMAN     STANDARD;      PRT;   312 AA.
AC   P02750;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85166241.
RA   TAKAHASHI N., TAKAHASHI Y., PUTNAM F.W.;
RT   "Periodicity of leucine and tandem repetition of a 24-amino acid
RT   segment in the primary structure of leucine-rich alpha 2-glycoprotein
RT   of human serum.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:1906-1910(1985).
CC   -!- FUNCTION: THE FUNCTION OF THIS PLASMA PROTEIN IS NOT KNOWN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS.
DR   PIR; A03211; NBHUA2.
DR   SWISS-2DPAGE; P02750; HUMAN.
DR   PFAM; PF00560; LRR; 4.
KW   PLASMA; GLYCOPROTEIN; REPEAT; LEUCINE-REPEAT.
FT   DISULFID      8     21
FT   DISULFID    268    294
FT   CARBOHYD      2      2
FT   CARBOHYD     44     44
FT   CARBOHYD    151    151
FT   CARBOHYD    234    234
FT   CARBOHYD    290    290
FT   CARBOHYD    271    271       POTENTIAL.
SQ   SEQUENCE   312 AA;  34346 MW;  48C3DB08 CRC32;

  Query Match           14.5%;  Score 144;  DB 1;  Length 312;
  Best Local Similarity  32.7%;  Pred. No. 1.50e-09;
  Matches   37;  Conservative   27;  Mismatches  43;  Indels   6;  Gaps   5;

Db    158 LDLGENQLETLPPDLLRGPLQ-LERLHLEGNKLQVLGKDLLLPQPDLRYLFLNGNKLARV 216
          | |  |||||:    :|| |  |  | | :| :  :::| :   : :| | |  |::: :
Qy     13 LMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTI 72

Db    217 AAGAFQGLRQLDMLDLSNNSLASVPEGLWASLGQPNWDMRDGFDISGNPWICD 269
          ::|||  |  |  ::| :|::      | |:||  :| :|    :||||: |:
Qy     73 TPGAFTTLVSLSTINLLSNPFNCNCH-LGAGLG--KW-LRKRRIVSGNPR-CQ 120


Search completed: Fri May 28 08:49:24 1999
Job time : 21 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:49:42 1999; MasPar time 11.79 Seconds
638.746 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-4
Description: (1-138) from US09191647.pep
Perfect Score: 995
Sequence: 1 EGAFNGAASVQELMLTGNQL.....CQKPFFLKEIPIQGVGHPGI 138

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 41.993; Variance 81.660; scale 0.514

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	857	86.1	1523	11	O88280	MEGF5.	1.46e-151
2	680	68.3	1531	11	O88279	MEGF4.	4.27e-114
3	233	23.4	739	4	O75094	MEGF5 (FRAGMENT).	8.60e-24
4	227	22.8	880	5	P91643	KEK1 PRECURSOR.	1.10e-22
5	223	22.4	184	4	O60803	DJ63G5.3 (PUTATIVE LEU	6.01e-22
6	215	21.6	733	5	Q24250	TARTAN PROTEIN PRECURS	1.75e-20
7	209	21.0	1496	4	Q92626	MYELOBLAST KIAA0230 (F	2.15e-19
8	208	20.9	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	3.26e-19
9	199	20.0	358	11	O70210	CHONDROADHERIN.	1.36e-17
10	196	19.7	358	11	O55226	CHONDROADHERIN.	4.65e-17
11	196	19.7	892	5	P91644	KEK2 PRECURSOR (FRAGME	4.65e-17
12	196	19.7	4293	11	O08852	POLYCYSTIC KIDNEY DISE	4.65e-17
13	195	19.6	907	4	O75473	ORPHAN G PROTEIN-COUPL	7.01e-17
14	193	19.4	359	4	O15335	CHONDROADHERIN.	1.59e-16
15	192	19.3	428	4	O14498	ISLR PRECURSOR.	2.39e-16
16	191	19.2	811	4	O75139	KIAA0644 PROTEIN.	3.60e-16
17	190	19.1	331	13	O93233	PHOSPHOLIPASE A2 INHIB	5.41e-16
18	179	18.0	603	11	O70211	INSULIN-LIKE GROWTH FA	4.61e-14
19	175	17.6	4292	4	Q15141	POLYCYSTIC KIDNEY DISE	2.28e-13
20	175	17.6	4302	4	Q15140	POLYCYSTIC KIDNEY DISE	2.28e-13
21	173	17.4	1066	5	Q18902	CODED FOR BY C. ELEGAN	5.04e-13
22	173	17.4	1385	5	Q26388	TLR=TOLL-LIKE RECEPTOR	5.04e-13
23	173	17.4	1389	5	Q24591	WHEELER.	5.04e-13
24	172	17.3	707	11	P97860	LEUCINE-RICH REPEAT PR	7.49e-13
25	171	17.2	677	6	Q28256	GLYCOPROTEIN IB.	1.11e-12
26	171	17.2	716	11	Q61809	LEUCINE-RICH-REPEAT PR	1.11e-12
27	169	17.0	3638	4	Q15142	POLYCYSTIC KIDNEY DISE	2.45e-12
28	168	16.9	610	5	Q21604	M88.6 PROTEIN.	3.64e-12
29	167	16.8	653	5	O02329	T23G11.6 PROTEIN.	5.39e-12
30	165	16.6	420	4	Q13641	5T4 ONCOFETAL ANTIGEN	1.18e-11
31	165	16.6	680	5	Q93374	C44H4.3 PROTEIN.	1.18e-11
32	165	16.6	705	4	O43377	BAC CLONE RG118D07 FRO	1.18e-11
33	163	16.4	713	4	O75325	GLIOMA AMPLIFIED ON CH	2.58e-11
34	162	16.3	313	4	Q13288	P37NB.	3.80e-11
35	159	16.0	738	5	Q93373	C44H4.2 PROTEIN.	1.22e-10
36	155	15.6	784	4	O60603	TOLL/INTERLEUKIN-1 REC	5.68e-10
37	155	15.6	784	4	O15454	TOLL-LIKE RECEPTOR 2.	5.68e-10
38	153	15.4	789	5	O16781	SIMILARITY TO MULTIPLE	1.22e-09
39	153	15.4	1355	5	O16779	CODED FOR BY C. ELEGAN	1.22e-09
40	152	15.3	658	10	O48788	PUTATIVE RECEPTOR KINA	1.79e-09
41	151	15.2	661	4	Q99467	LYMPHOCYTE ANTIGEN 64	2.62e-09
42	151	15.2	961	5	P90920	K07A12.2 PROTEIN.	2.62e-09
43	150	15.1	603	5	Q22075	T01G9.3 PROTEIN.	3.83e-09
44	150	15.1	1535	5	Q23991	PEROXIDASIN PRECURSOR.	3.83e-09
45	149	15.0	718	13	O73675	NEURONAL LEUCINE-RICH	5.59e-09

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 86.1%; Score 857; DB 11; Length 1523;
Best Local Similarity 89.6%; Pred. No. 1.46e-151;
Matches 120; Conservative 9; Mismatches 3; Indels 2; Gaps 2;

```
Db   574 EGAFDGAAGVQELMLTGNQLETMHGRMFRG-LSGLKTLMLRSNLISCVNNDTFAGLSSVR 632
             ||||:|||:|||||||||||||:||| ||| |||||||||||||:||:||||||||||
Qy     1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db   633 LLSLYDNRITTISPGAFTTLVSLSTINLLSNPFNCNCHM-AWLGRWLRKRRIVSGNPRCQ 691
             ||||||||||||:|||||||||||||||||||||||: | ||:|||||||||||||||
Qy    61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQ 120

Db   692 KPFFLKEIPIQDVA 705
             |||||||||| |:
Qy   121 KPFFLKEIPIQGVG 134
```

RESULT 2

```
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match           68.3%;  Score 680;  DB 11;  Length 1531;
  Best Local Similarity  65.9%;  Pred. No. 4.27e-114;
  Matches   91;  Conservative   29;  Mismatches  16;  Indels   2;  Gaps   2;

Db    582 DGTFEGATSVSELHLTANQLESVRSGMFRG-LDGLRTLMLRNNRISCIHNDSFTGLRNVR 640
          :|:|:||:|| || ||:||||:|::  ||| | ||:|||||:| |:|: ||:|:|| :||
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    641 LLSLYDNHITTISPGAFDTLQALSTLNLLANPFNCNCQL-AWLGDWLRKRKIVTGNPRCQ 699
          |||||||:||||:|||| || :|||:|||:|||||||:| | || |||||:||:||||||
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQ 120

Db    700 NPDFLRQIPLQDVAFPDF 717
          :| ||::||:| |: | :
Qy    121 KPFFLKEIPIQGVGHPGI 138


RESULT   3
ID   O75094      PRELIMINARY;      PRT;   739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;

  Query Match           23.4%;  Score 233;  DB 4;  Length 739;
  Best Local Similarity  36.6%;  Pred. No. 8.60e-24;
  Matches   34;  Conservative   24;  Mismatches  34;  Indels   1;  Gaps   1;

Db     12 SNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSLSHLALG 71
          : :| | ||:|  | : |::  :| || |:|:|:|  | |:::  |:|  | ||| : |
Qy     30 GGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPGAFTTLVSLSTINLL 89

Db     72 TNPLHCDCSLRW-LSEWVKAGYKEPGIARCSSP 103
          :||::|:| |   |: |::     :| :||  |
Qy     90 SNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT   4
ID   P91643      PRELIMINARY;      PRT;   880 AA.
AC   P91643;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KEK1 PRECURSOR.
GN   KEK1.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R;
RX   MEDLINE; 97223506.
RA   MUSACCHIO M., PERRIMON N.;
RT   "The Drosophila kekkon genes: novel members of both the leucine-rich
RT   repeat and immunoglobulin superfamilies expressed in the CNS.";
RL   DEV. BIOL. 178:63-76(1996).
DR   EMBL; U42767; G1150977; -.
DR   FLYBASE; FBgn0015399; kek1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    880       KEK1.
SQ   SEQUENCE   880 AA;  94542 MW;  1FF90059 CRC32;

  Query Match           22.8%;  Score 227;  DB 5;  Length 880;
  Best Local Similarity  32.7%;  Pred. No. 1.10e-22;
  Matches   37;  Conservative   29;  Mismatches  44;  Indels   3;  Gaps   3;

Db    195 PSLRELTLASNHIHKIESQAF-GNTPSLHKLDLSHCDIQTISAQAFGGLQGLTLLRLNGN 253
          :|::|| |::|::  : :::| |  ::|  | |    |  :| ::|:|| :: || |  |
Qy      8 ASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDN 67

Db    254 KLSELLPKTIETLSRLHGIELHDNPWLCDCRLRDT-KLWLMKRNIPYPVAPVC 305
          ::: : | :: ||  |  |:|  ||  |:|:|      || || |  :  | |
Qy     68 RITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIV-SGNPRC 119


RESULT   5
ID   O60803      PRELIMINARY;      PRT;   184 AA.
AC   O60803;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   DJ63G5.3 (PUTATIVE LEUCINE RICH PROTEIN) (FRAGMENT).
GN   DJ63G5.3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LLOYD D.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z94160; E1296586; -.
FT   NON_TER       1      1
FT   NON_TER     184    184
SQ   SEQUENCE   184 AA;  20630 MW;  7F359903 CRC32;

  Query Match           22.4%;  Score 223;  DB 4;  Length 184;
  Best Local Similarity  36.4%;  Pred. No. 6.01e-22;
```

```
  Matches   39;  Conservative   26;  Mismatches  40;  Indels   2;  Gaps   2;

Db     53 DGAFLGQSSLQVLQLGYNKLSNLTEGMLRG-MSRLQFLFVQHNLIEVVTPTAFSECPSLI 111
          :||| | :|:| | |  | |  :    :|| :| |  | :: |||  |:  :|:  :|:
Qy      1 EGAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVR 60

Db    112 SIDLSSNRLSRLDGATFASLASLMVCELAGNPFNCECDLFGFLA-WL 157
           : |  ||:: :  ::|::| ||   :| :|||||:| | : |: ||
Qy     61 LLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWL 107


RESULT    6
ID   Q24250      PRELIMINARY;      PRT;   733 AA.
AC   Q24250;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TARTAN PROTEIN PRECURSOR.
GN   TRN.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 94074761.
RA   CHANG Z., PRICE B.D., BOCKHEIM S., BOEDIGHEIMER M.J., SMITH R.,
RA   LAUGHON A.;
RT   "Molecular and genetic characterization of the Drosophila tartan
RT   gene.";
RL   DEV. BIOL. 160:315-332(1993).
DR   EMBL; U02078; G408375; -.
DR   FLYBASE; FBgn0010452; trn.
DR   PFAM; PF00560; LRR; 6.
KW   SIGNAL.
FT   SIGNAL       1     24       POTENTIAL.
SQ   SEQUENCE   733 AA;  81319 MW;  DA426BB0 CRC32;

  Query Match           21.6%;  Score 215;  DB 5;  Length 733;
  Best Local Similarity  35.3%;  Pred. No. 1.75e-20;
  Matches   41;  Conservative   28;  Mismatches  43;  Indels   4;  Gaps   4;

Db    195 FQAMPSLAELFLGMNTLQSIQAGAFQD-LKGLTRLELKGASLRNISHDSFLGLQELRILD 253
          | : :|: || |  | |::::: :|:  | ||  | |::  :  :|:|:| ||  :|:|
Qy      4 FNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLS 63

Db    254 LSDNRLDRIPSVGLSKLVRLEQLSLGQNDFEVISE-GAFMGLKQL-KRLEVNGALR 307
          | |||:  | : ::: || |  ::|  | |:     || :| | | ||  |:|  |
Qy     64 LYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLG-KWLRKRRIVSGNPR 118


RESULT    7
ID   Q92626      PRELIMINARY;      PRT;  1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELOBLAST KIAA0230 (FRAGMENT).
GN   KIAA0230.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA RES. 3:321-329(1996).
DR   EMBL; D86983; D1013908; -.
DR   PFAM; PF00047; ig; 4.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00141; peroxidase; 1.
DR   PFAM; PF00560; LRR; 3.
FT   NON_TER      1      1
SQ   SEQUENCE   1496 AA;  167209 MW;  5731EE51 CRC32;

  Query Match           21.0%;  Score 209;  DB 4;  Length 1496;
  Best Local Similarity  33.7%;  Pred. No. 2.15e-19;
  Matches   33;  Conservative   27;  Mismatches  37;  Indels   1;  Gaps   1;

Db    121 GAFEDLENLKYLYLYKNEIQSIDRQAFKG-LASLEQLYLHFNQIETLDPDSFQHLPKLER 179
          |||:   ::  | |  |:::::  ::|:| |::|  | |: | |  :  |:|  |: :
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    180 LFLHNNRITHLVPGTFNHLESMKRLRLDSNTLHCDCEI 217
          | | :|||| : ||:|  | |:  : | || ::|:| :
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHL 99


RESULT    8
ID   P70193      PRELIMINARY;      PRT;  1091 AA.
AC   P70193;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96394313.
RA   SUZUKI Y., SATO N., TOHYAMA M., WANAKA A., TAKAGI T.;
RT   "cDNA cloning of a novel membrane glycoprotein that is expressed
RT   specifically in glial cells in the mouse brain LIG-1: A protein with
RT   leucine-rich repeats and immunoglobulin-like domains.";
RL   J. BIOL. CHEM. 271:22522-22527(1996).
DR   EMBL; D78572; D1012081; -.
DR   MGD; MGI:107935; IMG.
DR   PFAM; PF00047; ig; 3.
DR   PFAM; PF00560; LRR; 7.
KW   MEMBRANE.
SQ   SEQUENCE   1091 AA;  119283 MW;  C0F262F9 CRC32;

  Query Match           20.9%;  Score 208;  DB 11;  Length 1091;
  Best Local Similarity  33.1%;  Pred. No. 3.26e-19;
  Matches   45;  Conservative   30;  Mismatches  57;  Indels   4;  Gaps   4;

Db    350 EGAFKGLKSLRVLDLDHNEISGTIEDTSGAFTGLDNLSKLTLFGNKIKSVAKRAFSGLES 409
          ||||:|  |:: | |  |::  |:    |  ||  |  | | :| |  |:: :|:|| |
Qy      1 EGAFNGAASVQELMLTGNQLE-TVHG-RGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSS 58

Db    410 LEHLNLGENAIRSVQFDAFAKMKNLKELYISSESFLCDCQLKW-LPPWLMGRMLQAFVTA 468
          :  |:| :| | ::   ||: : :|  : : |::| |:|:|   |  ||  | : :   :
Qy     59 VRLLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVS-GNP 117

Db    469 TCAHPESLKGQSIFSV 484
          |  |  ||  :| :|
Qy    118 RCQKPFFLKEIPIQGV 133


RESULT    9
ID   O70210      PRELIMINARY;      PRT;   358 AA.
AC   O70210;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CHONDROADHERIN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98129774.
RA   SHEN Z., GANTCHEVA S., MANSSON B., HEINEGARD D., SOMMARIN Y.;
RT   "Chondroadherin expression changes in skeletal development.";
RL   BIOCHEM. J. 330:549-557(1998).
DR   EMBL; AF004953; G2947105; -.
SQ   SEQUENCE   358 AA;  40403 MW;  47F9BF88 CRC32;

  Query Match           20.0%;  Score 199;  DB 11;  Length 358;
  Best Local Similarity  30.1%;  Pred. No. 1.36e-17;
  Matches   34;  Conservative   33;  Mismatches  42;  Indels   4;  Gaps   4;

Db    220 VEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDAAFAGVTTLKHVHLENNRL 279
          |:|||:|: | | ::   :|::    | || | :  :   |: :|||:::::: : | :||:
Qy     10 VQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRI 69

Db    280 NQL-PSTFP-FDNLETLTLTNNPWKCTCQL-RGLRRWL-EAKTSRPDATCSSP 328
          : |::|  : :| |: | :|| :| |:|  || :||   :    :: |  |
Qy     70 TTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT  10
ID   O55226      PRELIMINARY;      PRT;   358 AA.
AC   O55226;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CHONDROADHERIN.
GN   CHAD.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98126439.
RA   LANDGREN C., BEIER D.R., FAESSLER R., HEINEGARD D., SOMMARIN Y.;
RT   "The mouse chondroadherin gene: characterization and chromosomal
RT   localization.";
RL   GENOMICS 47:84-91(1998).
DR   EMBL; U96626; G2843110; -.
DR   MGD; MGI:1096866; CHAD.
SQ   SEQUENCE   358 AA;  40348 MW;  18A0CDB3 CRC32;

  Query Match           19.7%;  Score 196;  DB 11;  Length 358;
  Best Local Similarity  29.2%;  Pred. No. 4.65e-17;
  Matches   33;  Conservative   34;  Mismatches  42;  Indels   4;  Gaps   4;

Db    220 VEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDAAFSGVTTLKHVHLDNNRL 279
          |:|||:|: | | ::   :|::    | || | :  :   |: :|:|:::::: : | :||:
Qy     10 VQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRI 69

Db    280 NQL-PSSFP-FDNLETLTLTNNPWKCTCQL-RGLRRWL-EAKASRPDATCSSP 328
          : |::|  : :| |: | :|| :| |:|  || :||   :    :: |  |
Qy     70 TTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122


RESULT  11
ID   P91644      PRELIMINARY;      PRT;   892 AA.
AC   P91644;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KEK2 PRECURSOR (FRAGMENT).
GN   KEK2.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R;
RX   MEDLINE; 97223506.
RA   MUSACCHIO M., PERRIMON N.;
RT   "The Drosophila kekkon genes: novel members of both the leucine-rich
RT   repeat and immunoglobulin superfamilies expressed in the CNS.";
RL   DEV. BIOL. 178:63-76(1996).
DR   EMBL; U42768; G1150979; -.
DR   FLYBASE; FBgn0015400; kek2.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   NON_TER      1     1
FT   SIGNAL       <1    17       POTENTIAL.
FT   CHAIN        18    892      KEK2.
SQ   SEQUENCE   892 AA;  97294 MW;  93DCE450 CRC32;

  Query Match           19.7%;  Score 196;  DB 5;  Length 892;
  Best Local Similarity  31.8%;  Pred. No. 4.65e-17;
  Matches   42;  Conservative   27;  Mismatches  59;  Indels   4;  Gaps   4;

Db    119 TFQDYSSLMRLSLSGNPIRELKTSAFRH-LSFLTTLELSNCQVERIENEAFVGMDNLEWL 177
          :|   :|:  | |:|| :  :   :||  || | || | :  :  : |::| |: ::  |
Qy      3 AFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLL 62

Db    178 RLDGNRIGFI-QGTHILPKSLHGISLHSNRWNCDCRL-LDIHFWLVNYNTPLAEEPKCME 235
          | |||  | |:    ||  |:| ||  ||:|:|   :  || :    :: :|:|
Qy     63 SLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRK-RRIVSGNPRCQK 121

Db    236 PARLKGQVIKSL 247
          | ||   | ::
Qy    122 PFFLKEIPIQGV 133


RESULT  12
ID   O08852      PRELIMINARY;      PRT;  4293 AA.
AC   O08852;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   POLYCYSTIC KIDNEY DISEASE 1 PROTEIN.
GN   PKD1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97262094.
RA   LOHNING C., NOWICKA U., FRISCHAUF A.M.;
RT   "The mouse homolog of PKD1: sequence analysis and alternative
RT   splicing.";
RL   MAMM. GENOME 8:307-311(1997).
DR   EMBL; U70209; G2138183; -.
DR   PFAM; PF00560; LRR; 1.
DR   PFAM; PF00801; PKD; 16.
SQ   SEQUENCE   4293 AA;  466545 MW;  10E37A8A CRC32;

  Query Match           19.7%;  Score 196;  DB 11;  Length 4293;
  Best Local Similarity  34.2%;  Pred. No. 4.65e-17;
  Matches   40;  Conservative   28;  Mismatches  43;  Indels   6;  Gaps   5;

Db     56 LQTL-GPSLRIP-ADATALDLSHNLLQTLDIGLLVNLSALVELDLSNNRISTLEEGVFAN 113
          |:|: | ::|   :   :| |  ||:  :    :  ||::  | | :|||:|:  | |:
Qy     20 LETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPGAFTT 79

Db    114 LFNLSEINLSGNPFECNCGL-AWLPRWAKEHQVHVVQSEATTCRGPIPLAGQPLLSI 169
          | :|| ||| :|||:||| | | | :| : :  ::| |    |: |: |   |: ::
Qy     80 LVSLSTINLLSNPFNCNCHLGAGLGKWLRKR--RIV-SGNPRCQKPFFLKEIPIQGV 133


RESULT  13
ID   O75473      PRELIMINARY;      PRT;   907 AA.
AC   O75473;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE   ORPHAN G PROTEIN-COUPLED RECEPTOR HG38.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98308104.
RA   MCDONALD T., WANG R., BAILEY W., XIE G., CHEN F., CASKEY C.T.,
RA   LIU Q.;
RT   "Identification and cloning of an orphan G protein-coupled receptor
RT   of the glycoprotein hormone receptor subfamily.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 247:266-270(1998).
DR   EMBL; AF062006; G3366802; -.
SQ   SEQUENCE   907 AA;  99997 MW;  B9147406 CRC32;

  Query Match           19.6%;  Score 195;  DB 4;  Length 907;
  Best Local Similarity  40.2%;  Pred. No. 7.01e-17;
  Matches   35;  Conservative   16;  Mismatches  35;  Indels   1;  Gaps   1;

Db    108 GAFTGLYSLKVLMLQNNQLRHVPTEALQN-LRSLQSLRLDANHISYVPPSCFSGLHSLRH 166
          ||| |  |:  |||  |||  |   :::  | :| :| | :| |: |:   |:|| |:|
Qy      2 GAFNGAASVQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRL 61

Db    167 LWLDDNALTEIPVQAFRSLSALQAMTL 193
          | | || :| |   || :| :| :: |
Qy     62 LSLYDNRITTITPGAFTTLVSLSTINL 88


RESULT   14
ID   O15335      PRELIMINARY;      PRT;   359 AA.
AC   O15335;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CHONDROADHERIN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98008928.
RA   GROVER J., CHEN X.N., KORENBERG J.R., ROUGHLEY P.J.;
RT   "The structure and chromosome location of the human chondroadherin
RT   gene (CHAD).";
RL   GENOMICS 45:379-385(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   GROVER J., ROUGHLEY P.J.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   GROVER J., ROUGHLEY P.J.;
RL   SUBMITTED (APR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U96769; G2104968; -.
DR   EMBL; U96767; G2104968; JOINED.
DR   EMBL; U96768; G2104968; JOINED.
DR   PFAM; PF00560; LRR; 5.
SQ   SEQUENCE   359 AA;  40487 MW;  22CBC94E CRC32;

  Query Match           19.4%;  Score 193;  DB 4;  Length 359;
  Best Local Similarity  30.1%;  Pred. No. 1.59e-16;
  Matches   34;  Conservative   31;  Mismatches  44;  Indels   4;  Gaps   4;

Db    221 VEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDGAFLGVTTLKHVHLENNRL 280
          |:||:|: | | ::   :|::    | || | :  :   |: :| |::::: : | :||:
Qy     10 VQELMLTGNQLETVHGRGFRGGLSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRI 69

Db    281 NQL-PSNFP-FDSLETLALTNNPWKCTCQL-RGLRRWL-EAKASRPDATCASP 329
            : |: |  : || |: | :|| :| |:|  || :||   :    :: |  |
Qy     70 TTITPGAFTTLVSLSTINLLSNPFNCNCHLGAGLGKWLRKRRIVSGNPRCQKP 122

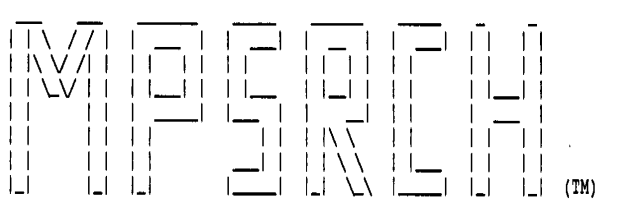

RESULT   15
ID   O14498      PRELIMINARY;      PRT;   428 AA.
AC   O14498;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ISLR PRECURSOR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 97468140.
RA   NAGASAWA A., KUBOTA R., IMAMURA Y., NAGAMINE K., WANG  Y., ASAKAWA S.,
RA   KUDOH J., MINOSHIMA S., MASHIMA Y., OGUCHI Y., SHIMIZU N.;
RT   "Cloning of the cDNA for a new member of the immunoglobulin
RT   superfamily (ISLR) containing leucine-rich repeat (LRR).";
RL   GENOMICS 44:273-279(1997).
DR   EMBL; AB003184; D1023718; -.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL       1     18       POTENTIAL.
SQ   SEQUENCE   428 AA;  45997 MW;  C83DFE0B CRC32;

  Query Match           19.3%;  Score 192;  DB 4;  Length 428;
  Best Local Similarity  35.3%;  Pred. No. 2.39e-16;
  Matches   36;  Conservative   20;  Mismatches  42;  Indels   4;  Gaps   4;

Db    121 LSALQLLKMDSNELTFIPRDAFRSLRALRSLQLNHNRLHTLAEGTFTPLTALSHLQINEN 180
          ||:|  |:: || :  :: |:| :| ::| | |  ||: |:: |:|| | :|| : :  |
Qy     32 LSGLKTLMLRSNLIGCVSNDTFAGLSSVRLLSLYDNRITTITPGAFTTLVSLSTINLLSN 91

Db    181 PFDCTCGI-VWLKTWALTTAVSIPEQDNIACTSPHVLKGTPL 221
          ||:| | :   |  | |     |    |  | |  ||  |:
Qy     92 PFNCNCHLGAGLGKW-LRKR-RIVS-GNPRCQKPFFLKEIPI 130


Search completed: Fri May 28 08:50:23 1999
Job time : 41 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:52:24 1999; MasPar time 8.51 Seconds
399.730 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-5
Description: (1-160) from US09191647.pep
Perfect Score: 1212
Sequence: 1 WPRCECMPGYAGDNCSENQD.....NGGNDHIAVXLYXGHVRFSY 160

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 30.075; Variance 148.234; scale 0.203

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1158	95.5	1534	30	W46966	Amino acid sequence o	1.40e-87
2	489	40.3	1480	5	R25079	Drosophila SLIT prote	1.58e-30
3	318	26.2	520	25	W18348	Proliferation and dif	2.11e-16
4	318	26.2	702	25	W18349	Proliferation and dif	2.11e-16
5	318	26.2	723	25	W18353	Proliferation and dif	2.11e-16
6	306	25.2	660	21	W11725	H-Delta-1 polypeptide	1.99e-15
7	304	25.1	727	21	W11719	C-Delta-1 polypeptide	2.89e-15
8	304	25.1	740	21	W00876	C-Delta-1 polypeptide	2.89e-15
9	302	24.9	722	21	W11720	M-Delta-1 polypeptide	4.19e-15
10	298	24.6	2321	36	W49698	Human Notch3 protein.	8.83e-15
11	291	24.0	1872	36	W68510	Partial human Notch-3	3.24e-14
12	284	23.4	1193	19	W05835	Chick Serrate.	1.19e-13
13	283	23.3	1036	25	W18351	Proliferation and dif	1.43e-13
14	283	23.3	1187	25	W18352	Proliferation and dif	1.43e-13
15	283	23.3	1218	29	W44301	Human serrate 1.	1.43e-13
16	283	23.3	1218	19	W05833	Human Serrate-1 (HJ1)	1.43e-13
17	283	23.3	1218	25	W18354	Proliferation and dif	1.43e-13
18	278	22.9	1208	28	W40827	Human Jagged protein.	3.61e-13
19	271	22.4	685	37	W80813	Nucleotide sequence o	1.31e-12
20	272	22.4	2707	24	W27161	Mouse receptor ME2.	1.09e-12
21	268	22.1	1055	29	W44298	Human serrate 2 prote	2.28e-12
22	268	22.1	1212	29	W44299	Human serrate 2.	2.28e-12
23	268	22.1	1257	19	W05834	Human Serrate-2 (HJ2)	2.28e-12
24	259	21.4	612	28	W39256	Human partial mature	1.19e-11
25	259	21.4	737	28	W39257	Human membrane protei	1.19e-11
26	241	19.9	1404	7	R38304	Sequence of a serrate	3.21e-10
27	236	19.5	157	21	W11730	H-Delta-1 polypeptide	7.98e-10
28	229	18.9	2409	3	R12609	Versican.	2.84e-09
29	227	18.7	833	6	R28960	Delta D11.	4.08e-09
30	225	18.6	1257	9	R46627	Neurocan core protein	5.86e-09
31	220	18.2	1572	24	W27160	Mouse receptor ME2 re	1.45e-08
32	218	18.0	385	10	R56167	Neuroendocrine tumor	2.07e-08
33	207	17.1	473	17	R86869	Adhesive protein.	1.49e-07
34	206	17.0	77	6	R28962	ELR-11 and -12.	1.79e-07
35	204	16.8	4544	11	R60517	Human alpha-2-MR.	2.55e-07
36	204	16.8	4544	9	R47861	Alpha 2-Macroglobulin	2.55e-07
37	201	16.6	383	10	R56166	Neuroendocrine tumor	4.36e-07
38	180	14.9	2189	1	R05222	Antigen GX5401FL enco	1.79e-05
39	178	14.7	228	30	W46967	Amino acid sequence o	2.54e-05
40	171	14.1	221	21	W10366	Murine Del-1 truncate	8.61e-05
41	171	14.1	480	21	W10364	Mouse developmentally	8.61e-05
42	166	13.7	481	21	W10365	Human developmentally	2.05e-04
43	161	13.3	810	27	W37500	Human nel-related pro	4.86e-04
44	158	13.0	912	14	R85442	Bovine brevican core	8.14e-04
45	157	13.0	2476	39	W67738	Pig p105 zona pelluci	9.67e-04

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 95.5%; Score 1158; DB 30; Length 1534;
Best Local Similarity 95.6%; Pred. No. 1.40e-87;
Matches 152; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1067 prcecmpgyagdncsenqddcrdhrcqngaqcmdevnsysclcaegysgqlceipphlpa 1126
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2 PRCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA 61

Db 1127 pkspcegtecqngancvdqgnrpvcqclpgfggpecekllsvnfvdrdtylqftdlqnwp 1186
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Qy     62 PKSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWX 121

Db   1187 ranitlqvstaedngillyngdndhiavelyqghvrvsy 1225
          | |||||| |||||||||||| |||||| || |||| ||
Qy    122 RXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match          40.3%;  Score 489;  DB 5;  Length 1480;
  Best Local Similarity  40.5%;  Pred. No. 1.58e-30;
  Matches   66;  Conservative   42;  Mismatches  48;  Indels   7;  Gaps   5;

Db   1050 cdcqagfhgtnctdniddcqnhmcqnggtcvdgindyqcrcpddytgkyceghnmismmy 1109
           |:| :|: | ||::| |||::| ||||: |:| :| | | |:: |:|  ||    ::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db   1110 pqtspcqnheckhgvcfqpnaqgsdylcrchpgytgkwceyltsisfvhnnsfveleplr 1169
           |  |||:  || :|      ||:  :|:| ||: |  || | |::||  :::::  |:
Qy     62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db   1170 trpeanvti-vfssaeqngilmydgqdahlavelfngrirvsy 1211
            :   |:|: ||: ||:||||:|:| : |:|| |: |::| ||
Qy    119 NWXRXNITLQVFT-AEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT    3
ID   W18348 standard; protein; 520 AA.
AC   W18348;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 3; Page 59-61; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   520 AA;

  Query Match          26.2%;  Score 318;  DB 25;  Length 520;
  Best Local Similarity  43.4%;  Pred. No. 2.11e-16;
  Matches   43;  Conservative   24;  Mismatches  26;  Indels   6;  Gaps   5;

Db    408 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-ap-v-s-r 463
          | | :|::| :| :| |||    | ||: | | || :|| |: ||:|: |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    464 --cehapchngatcherghryvcecargyggpncqfllp 500
              || : |:||| | ::|:| ||:|  |:|||:|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    4
ID   W18349 standard; protein; 702 AA.
AC   W18349;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
```

PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 4; Page 61-64; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 702 AA;

Query Match 26.2%; Score 318; DB 25; Length 702;
Best Local Similarity 43.4%; Pred. No. 2.11e-16;
Matches 43; Conservative 24; Mismatches 26; Indels 6; Gaps 5;

Db 408 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-ap-v-s-r 463
Qy 4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db 464 --cehapchngatcherghryvcecargyggpncqfllp 500
Qy 64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102

RESULT 5
ID W18353 standard; protein; 723 AA.
AC W18353;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;
KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW immunosuppression.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal
FT Protein 22..723
FT /label= Differentiation_suppression_protein
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
DR N-PSDB; T70174.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 15; Page 77-82; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 723 AA;

Query Match 26.2%; Score 318; DB 25; Length 723;
Best Local Similarity 43.4%; Pred. No. 2.11e-16;
Matches 43; Conservative 24; Mismatches 26; Indels 6; Gaps 5;

Db 429 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-ap-v-s-r 484
Qy 4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db 485 --cehapchngatcherghryvcecargyggpncqfllp 521
Qy 64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102

RESULT 6
ID W11725 standard; Protein; 660 AA.
AC W11725;
DT 28-APR-1997 (first entry)
DE H-Delta-1 polypeptide (reading frame 1 product).
KW H-Delta-1; cell proliferation; nervous system disorder;
KW tissue regeneration; Notch; cervix cancer; breast cancer;
KW lung cancer; colon cancer; melanoma; seminoma;
KW neurogenesis; therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 219..221
FT /note= "Delta-1 homologous region"
FT region 245..246
FT /note= "Delta-1 homologous region"
FT region 259..428
FT /note= "Delta-1 homologous region"
FT region 430..434
FT /note= "Delta-1 homologous region"
FT region 594..597
FT /note= "Delta-1 homologous region"
FT region 605..608
FT /note= "Delta-1 homologous region"
FT region 615..617
FT /note= "Delta-1 homologous region"
FT misc_difference 32
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 40
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 41
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 87
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 138
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 139
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 145
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 162
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 187
FT /note= "undetermined amino acid residue"
FT misc_difference 203
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 230
FT /note= "undetermined amino acid residue"
FT misc_difference 249
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 429
FT /note= "undetermined amino acid residue"
FT misc_difference 447
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 498

```
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 513
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 541
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 552
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 556
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 564
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 576
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 580
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 619
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 621
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 626
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 630
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 634
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 639
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 642
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 643
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 644
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 647
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 648
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 651
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 652
FT                   /note= "undetermined amino acid residue"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T59454.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 12B1-6; 135pp; English.
CC   Polypeptide sequences (W11725-27) were determined for all 3
CC   reading frames of a human H-delta-1 contig sequence (T59454) obtd.
CC   from a foetal brain library.  Errors in the contig sequence meant
CC   that no single reading frame gave the correct sequence for the
CC   H-Delta-1 protein.  The 3 polypeptide sequences were therefore
CC   compared to chick and mouse Delta-1 sequences (see also W11719-20)
CC   and regions of homology (see also W11728-38) identified.  H-Delta-1
CC   is the human homologue of Drosophila Delta, a protein that binds
CC   to Notch protein.  H-Delta-1 polypeptides can be used to treat
CC   disorders of cell fate or differentiation, such as cancer, and
CC   nervous system disorders, or to promote tissue regeneration and
CC   repair.
SQ   Sequence   660 AA;

  Query Match          25.2%;  Score 306;  DB 21;  Length 660;
  Best Local Similarity  42.4%;  Pred. No. 1.99e-15;
  Matches   42;  Conservative   24;  Mismatches  27;  Indels   6;  Gaps   3;

Db    341 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrnc--sa--pasr 396
          | | :|::| :| :| |||    | ||: | | || :|| |: ||:|: |  ::  ||::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    397 --cehapchngatcherghryxcecarsyggpncxfllp 433
            || : |:||| | ::|:|  |:|  ::|||:|  ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT   7
ID   W11719 standard; Protein; 727 AA.
AC   W11719;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide.
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          184..228
FT                   /label= DSL
FT   domain          229..261
FT                   /label= EGF1
FT   domain          262..292
FT                   /label= EGF2
FT   domain          293..332
FT                   /label= EGF3
FT   domain          333..370
FT                   /label= EGF4
FT   domain          371..409
FT                   /label= EGF5
FT   domain          410..447
FT                   /label= EGF6
FT   domain          448..485
FT                   /label= EGF7
FT   domain          486..523
FT                   /label= EGF8
FT   domain          524..534
FT                   /label= EGF9
FT   domain          555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58897.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W11719) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
CC   C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
CC   amino acid sequence was deduced from a cDNA clone (T58897) obtd.
CC   from chick stage 4-6 embryos.  An alternatively spliced variant
CC   (W00876) was also isolated, and mouse (W11720) and human (W11721-
CC   38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC   can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
```

```
CC   cancer, melanoma or seminoma, and nervous system disorders or to
CC   promote tissue regeneration and repair.
SQ   Sequence   727 AA;

  Query Match          25.1%;  Score 304;  DB 21;  Length 727;
  Best Local Similarity  43.4%;  Pred. No. 2.89e-15;
  Matches   43;  Conservative   21;  Mismatches  29;  Indels   6;  Gaps   4;

Db    436 cqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcppgyngkncstp--v-s-r 491
          |:| :|::| :| :| |||    | ||: | | || ||| |: ||:|  |  |  : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    492 --cehnpchngatchersnryvcecargygglncqfllp 528
            ||   |:||| | :::|| ||:|  |:|| :|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


SULT    8
    W00876 standard; Protein; 740 AA.
AC  W00876;
DT  28-APR-1997  (first entry)
DE  C-Delta-1 polypeptide (alternatively spliced variant).
KW  C-Delta-1; cell proliferation; nervous system disorder;
KW  tissue regeneration; Notch; cervix cancer; breast cancer;
KW  lung cancer; colon cancer; melanoma; seminoma;
KW  neurogenesis; therapy.
OS  Gallus sp.
FH  Key             Location/Qualifiers
FT  domain          184..228
FT                  /label= DSL
FT  domain          229..261
FT                  /label= EGF1
FT  domain          262..292
FT                  /label= EGF2
FT  domain          293..332
FT                  /label= EGF3
FT  domain          333..370
FT                  /label= EGF4
FT  domain          371..409
FT                  /label= EGF5
FT  domain          410..447
FT                  /label= EGF6
FT  domain          448..485
FT                  /label= EGF7
FT  domain          486..523
FT                  /label= EGF8
    domain          524..534
FT                  /label= EGF9
FT  domain          555..579
FT                  /label= TM
FT                  /note= "transmembrane domain"
PN  WO9701571-A1.
PD  16-JAN-1997.
PF  28-JUN-1996; U11178.
PR  28-JUN-1995; US-000589.
PA  (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA  (UYYA ) UNIV YALE.
PI  Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI  Lewis J;
DR  WPI; 97-100159/09.
DR  N-PSDB; T58898.
PT  New vertebrate Delta protein, DNA and antibodies - for treating and
PT  preventing cancer, nervous system disorders and for tissue
PT  regeneration
PS  Disclosure; Fig 2; 135pp; English.
CC  C-delta-1 polypeptide (W00876) is the chick homologue of Drosophila
CC  Delta, a protein that binds to Notch protein.  Expression of
CC  C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
CC  amino acid sequence was deduced from a cDNA clone (T58898) obtd.
CC  from chick stage 4-6 embryos.  A shorter version (W58877) of
CC  C-Delta-1, lacking the 12 C-terminal amino acids of the longer
CC  version, was also isolated, and mouse (W11720) and human (W11721-
CC  38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC  can be used to treat or prevent disorders characterised by
CC  increased Notch activity, such as cervical, breast, lung or colon
CC  cancer, melanoma or seminoma, and nervous system disorders or to
CC  promote tissue regeneration and repair.
SQ  Sequence   740 AA;

  Query Match          25.1%;  Score 304;  DB 21;  Length 740;
  Best Local Similarity  43.4%;  Pred. No. 2.89e-15;
  Matches   43;  Conservative   21;  Mismatches  29;  Indels   6;  Gaps   4;

Db    436 cqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcppgyngkncstp--v-s-r 491
          |:| :|::| :| :| |||    | ||: | | || ||| |: ||:|  |  |  : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    492 --cehnpchngatchersnryvcecargygglncqfllp 528
            ||   |:||| | :::|| ||:|  |:|| :|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    9
ID  W11720 standard; Protein; 722 AA.
AC  W11720;
DT  28-APR-1997  (first entry)
DE  M-Delta-1 polypeptide.
KW  M-Delta-1; cell proliferation; nervous system disorder;
KW  tissue regeneration; Notch; cervix cancer; breast cancer;
KW  lung cancer; colon cancer; melanoma; seminoma;
KW  neurogenesis; therapy.
OS  Mus sp.
PN  WO9701571-A1.
PD  16-JAN-1997.
PF  28-JUN-1996; U11178.
PR  28-JUN-1995; US-000589.
PA  (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA  (UYYA ) UNIV YALE.
PI  Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI  Lewis J;
DR  WPI; 97-100159/09.
DR  N-PSDB; T58899.
PT  New vertebrate Delta protein, DNA and antibodies - for treating and
PT  preventing cancer, nervous system disorders and for tissue
PT  regeneration
PS  Claim 4; Fig 8; 135pp; English.
CC  M-delta-1 polypeptide (W11720) is the mouse homologue of Drosophila
CC  Delta, a protein that binds to Notch protein.  It is expressed
CC  primarily in presomitic mesoderm, the central and peripheral
CC  nervous systems, and kidney.  Chick (W11719) and human (W11721-
CC  38) Delta-1 polypeptides have also been identified.  Delta-1
CC  proteins can be used to treat or prevent disorders characterised by
CC  increased Notch activity, such as cervical, breast, lung or colon
CC  cancer, melanoma or seminoma, as well as nervous system disorders,
CC  and to promote tissue regeneration and repair.
SQ  Sequence   722 AA;

  Query Match          24.9%;  Score 302;  DB 21;  Length 722;
  Best Local Similarity  42.4%;  Pred. No. 4.19e-15;
  Matches   42;  Conservative   22;  Mismatches  29;  Indels   6;  Gaps   5;

Db    428 crcqagfsgrycedndddcasspcanggtcrdsvndfsctcppgytgkncs-ap-v-s-r 483
          | | :|::|  | :| |||    | ||: | | || :|| |: ||:|  |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    484 --cehapchngatchqrgqrymcecaqgyggpncqfllp 520
            || : |:||| | ::| | :|:|  |:|||:|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT   10
ID  W49698 standard; Protein; 2321 AA.
AC  W49698;
DT  21-DEC-1998  (first entry)
DE  Human Notch3 protein.
KW  Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
```

KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leukoencephalopathy; therapy.
OS Homo sapiens.
PN FR2751986-A1.
PD 06-FEB-1998.
PF 16-APR-1997; 004680.
PR 01-AUG-1996; FR-009733.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR WPI; 98-133138/13.
DR N-PSDB; V57001.
PT Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leukoencephalopathy
PS Claim 2; Fig 1.1-1.8; 45pp; French.
CC This sequence represents the human Notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of
CC the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC and leukoencephalopathy (CADASIL) type. Blocking expression of a
CC mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC gene or protein can be used to treat CADASIL or related disorders.
SQ Sequence 2321 AA;

Query Match 24.6%; Score 298; DB 36; Length 2321;
Best Local Similarity 42.7%; Pred. No. 8.83e-15;
Matches 41; Conservative 15; Mismatches 39; Indels 1; Gaps 1;

```
Db   1108 ceclpgyngdnceddvdecasqpcqhggscidlvarylcscppgtlgvlceineddcgpg 1167
          |||:||| |||| :: |:|  : ||:|: |:| |  | | |: |  | ||||     :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1168 ppldsgprclhngtcvdlvggfrctcppgytglrce 1203
          :| : |  | : : |||      | | ||: |  ||
Qy     64 SPCE-GTECQNGANCVDQGNRPVCQCLPGFGGPECE 98
```

RESULT 11
ID W68510 standard; Protein; 1872 AA.
AC W68510;
DT 06-JAN-1999 (first entry)
DE Partial human Notch-3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leukoencephalopathy.
OS Homo sapiens.

	Key	Location/Qualifiers
FT	Misc_difference	328
FT		/note= "encoded by NAN"
FT	Misc_difference	401
FT		/note= "encoded by GNN"
FT	Misc_difference	403
FT		/note= "encoded by GNC"
FT	Misc_difference	406
FT		/note= "encoded by GNN"
FT	Misc_difference	409
FT		/note= "encoded by NNT"
FT	Misc_difference	420
FT		/note= "encoded by GNC"
FT	Misc_difference	706
FT		/note= "encoded by NNN"
FT	Misc_difference	708
FT		/note= "encoded by CCN"
FT	Misc_difference	719
FT		/note= "encoded by CGN"
FT	Misc_difference	728
FT		/note= "encoded by CNT"
FT	Misc_difference	729
FT		/note= "encoded by GTN"
FT	Misc_difference	759..789
FT		/note= "encoded by NNN"
FT	Misc_difference	1425
FT		/note= "encoded by GNA"

PN FR2751985-A1.
PD 06-FEB-1998.
PF 01-AUG-1996; 009733.
PR 01-AUG-1996; FR-009733.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR WPI; 98-133137/13.
DR N-PSDB; V57163.
PT Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leukoencephalopathy
PS Claim 2; Fig 1a-1g; 42pp; French.
CC This sequence represents a partial human notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of the
CC cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) type. Blocking expression of a mutated
CC Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC protein can be used to treat CADASIL or related disorders.
SQ Sequence 1872 AA;

Query Match 24.0%; Score 291; DB 36; Length 1872;
Best Local Similarity 41.7%; Pred. No. 3.24e-14;
Matches 40; Conservative 16; Mismatches 39; Indels 1; Gaps 1;

```
Db   1042 ceclpgyngdnceddvdecasqpcqhggscidlvarylcscppgtlgvfceineddcgpg 1101
          |||:||| |||| :: |:|  : ||:|: |:| |  | | |: |  | :|||     :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1102 ppldsgprclhngtcvdlvggfrctcppgytglrce 1137
          :| : |  | : : |||      | | ||: |  ||
Qy     64 SPCE-GTECQNGANCVDQGNRPVCQCLPGFGGPECE 98
```

RESULT 12
ID W05835 standard; Protein; 1193 AA.
AC W05835;
DT 28-JAN-1997 (first entry)
DE Chick Serrate.
KW C-Serrate; Notch; cell differentiation; cell fate; tissue repair;
KW central nervous system; cancer; therapy; diagnosis.
OS Gallus sp.

FH	Key	Location/Qualifiers
FT	domain	1..1041
FT		/label= Extracellular_domain
FT	peptide	1..5
FT		/label= Sig_peptide
FT		/note= "lacks the N-terminal portion owing to
FT		truncation of the encoding cDNA clone"
FT	domain	158..203
FT		/label= DSL
FT		/note= "region of homology with Drosophila Delta
FT		and Serrate, predicted to mediate binding
FT		with Notch"
FT	domain	208..837
FT		/label= ELR
FT		/note= "epidermal growth factor-like repeat domain"
FT	region	208..238
FT		/label= ELR1
FT	region	239..274
FT		/label= ELR2
FT	region	275..313
FT		/label= ELR3
FT	region	314..351
FT		/label= ELR4
FT	region	352..390
FT		/label= ELR5
FT	region	391..427
FT		/label= ELR6

```
FT   region          428..464
FT                   /label= ELR7
FT   region          465..502
FT                   /label= ELR8
FT   region          503..540
FT                   /label= ELR9
FT   region          541..606
FT                   /label= ELR10
FT   region          607..644
FT                   /label= ELR11
FT   region          655..682
FT                   /label= ELR12
FT   region          683..721
FT                   /label= ELR13
FT   region          722..759
FT                   /label= ELR14
FT   region          760..797
FT                   /label= ELR15
FT   region          798..837
FT                   /label= ELR16
FT   region          854..911
FT                   /label= Cysteine-rich_region
FT   domain          1042..1066
FT                   /label= Transmembrane_domain
FT   domain          1067..1193
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI   Lewis JH, Mann RS, Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40092.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Disclosure; Page 112-115; 161pp; English.
CC   Chicken Serrate (W05835), or C-Serrate, is a ligand for the zygotic
CC   neurogenic locus Notch and is believed to play a major role in
CC   determining cell fates in the central nervous system. Its amino
CC   acid sequence was deduced from a cDNA clone (T40092) obtd. from an
CC   optic explant cDNA library. C-Serrate is expressed in the central
CC   nervous system, cranial placodes, nephric mesoderm, vascular
CC   system, and limb bud mesenchyme.
SQ   Sequence   1193 AA;

  Query Match          23.4%;  Score 284;  DB 19;  Length 1193;
  Best Local Similarity  40.8%;  Pred. No. 1.19e-13;
  Matches   40;  Conservative   17;  Mismatches  35;  Indels   6;  Gaps   3;

Db    446 rcicspgyagdhcekdinecasnpcmngghcqdeingfqclcpagfsgnlc----qldi- 500
          || | ||||||:|  : ::|  : | ||::| ||:|:: |||: |:|| ||    :|
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    501 dy-cepnpcqngaqcfnlamdyfcncpedyegkncshl 537
           ||   ||||| | : :    | |   : | :|  |
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKL 100


RESULT   13
ID   W18351 standard; protein; 1036 AA.
AC   W18351;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 5; Page 66-71; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1036 AA;

  Query Match          23.3%;  Score 283;  DB 25;  Length 1036;
  Best Local Similarity  40.8%;  Pred. No. 1.43e-13;
  Matches   40;  Conservative   18;  Mismatches  34;  Indels   6;  Gaps   3;

Db    441 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlc----qldi- 495
          || | ||||||:|  : |:|  : | ||::| :|:| : |||: |:|| ||    :|
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    496 dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 532
           ||   ||||| | ::::   | |   : | :|  |
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKL 100


RESULT   14
ID   W18352 standard; protein; 1187 AA.
AC   W18352;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 6; Page 71-76; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1187 AA;

  Query Match          23.3%;  Score 283;  DB 25;  Length 1187;
  Best Local Similarity  40.8%;  Pred. No. 1.43e-13;
  Matches   40;  Conservative   18;  Mismatches  34;  Indels   6;  Gaps   3;

Db    441 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlc----qldi- 495
          || | ||||||:|  : |:|  : | ||::| :|:| : |||: |:|| ||    :|
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    496 dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 532
           ||   ||||| | ::::   | |   : | :|  |
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKL 100


RESULT   15
```

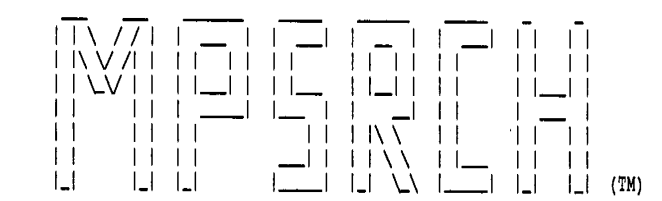
ID W44301 standard; Protein; 1218 AA.
AC W44301;
DT 19-JUN-1998 (first entry)
DE Human serrate 1.
KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= Signal
FT Protein 32..1218
FT /label= Serrate-1
PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15201.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Disclosure; Page 77-86; 103pp; Japanese.
CC The present sequence represents human serrate 1, from the present
CC invention which describes human serrate 2. The present invention also
CC describes a method for the preparation of the polypeptides, and
CC antibodies binding to the polypeptide and its fragments. The polypeptide
CC and its fragments expressed by the serrate-2-gene can be used to inhibit
CC stem (especially blood stem) cell differentiation and to inhibit
CC endothelial cell growth. They may be incorporated in a cell culture
CC media for culturing undifferentiated stem cells. They can also be used
CC for treatment of neoplasms such as leukaemia. The antibodies can be used
CC for the diagnosis of malignant tumours.
SQ Sequence 1218 AA;

Query Match 23.3%; Score 283; DB 29; Length 1218;
Best Local Similarity 40.8%; Pred. No. 1.43e-13;
Matches 40; Conservative 18; Mismatches 34; Indels 6; Gaps 3;

```
Db   472 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlc----qldi- 526
         || | ||||||:|  : |:|  : | ||::| :|:| : |||: |:|| ||    :|
Qy     3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db   527 dy-cepnpcqngaqcynrasdyfckcpedyegkncshl 563
           ||   ||||| | ::::   | |   : | :|  |
Qy    63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKL 100
```

rch completed: Fri May 28 08:53:26 1999
time : 62 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:53:43 1999; MasPar time 9.24 Seconds
693.556 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-5
Description: (1-160) from US09191647.pep
Perfect Score: 1212
Sequence: 1 WPRCECMPGYAGDNCSENQD.....NGGNDHIAVXLYXGHVRFSY 160

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 40.603; Variance 78.520; scale 0.517

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	491	40.5	530	2	A31640	epidermal growth fact	3.64e-76
2	489	40.3	1469	2	B36665	slit protein 2 precur	9.43e-76
3	489	40.3	1480	2	A36665	slit protein 1 precur	9.43e-76
4	343	28.3	293	2	B26637	neurogenic repetitive	5.00e-46
5	343	28.3	2139	2	A35672	crumbs protein - frui	5.00e-46
6	308	25.4	1064	2	A40136	fibropellin Ia - sea	4.27e-39
7	304	25.1	728	2	I50719	C-Delta-1 - chicken	2.61e-38
8	302	24.9	722	2	I48324	DELTA-like 1 - mouse	6.44e-38
9	298	24.6	2321	2	S78549	notch3 protein - huma	3.91e-37
10	294	24.3	1203	2	A49175	Motch B protein - mou	2.36e-36
11	291	24.0	2437	2	S42612	transmembrane protein	9.10e-36
12	287	23.7	570	2	A48836	fibropellin C precurs	5.47e-35
13	287	23.7	2471	2	A49128	cell-fate determining	5.47e-35
14	283	23.3	1220	2	A56136	jagged protein precur	3.28e-34
15	279	23.0	2318	2	S45306	notch 3 protein - mou	1.95e-33
16	273	22.5	387	2	B49175	Motch A protein - mou	2.83e-32
17	273	22.5	2531	2	A46019	gene Notch-1 protein	2.83e-32
18	272	22.4	2531	2	S18188	notch protein homolog	4.41e-32
19	272	22.4	2555	2	A40043	notch protein homolog	4.41e-32
20	269	22.2	2703	2	A24420	notch protein - fruit	1.67e-31
21	268	22.1	2524	2	A35844	Xotch protein - Afric	2.60e-31
22	260	21.5	861	2	A48825	Notch homolog Motch p	8.92e-30
23	246	20.3	1429	2	S06434	homeotic protein lin-	4.15e-27
24	241	19.9	1404	2	A36666	serrate protein precu	3.67e-26
25	241	19.9	1408	2	S16148	gene serrate protein	3.67e-26
26	231	19.1	3562	2	A47171	chondroitin sulfate p	2.79e-24
27	229	18.9	102	2	B55885	chondroitin sulfate p	6.61e-24
28	229	18.9	2409	2	A60979	versican precursor -	6.61e-24
29	227	18.7	832	2	A31246	neurogenic protein De	1.56e-23
30	227	18.7	833	2	S19087	gene Delta protein pr	1.56e-23
31	227	18.7	880	2	S00670	gene Delta protein pr	1.56e-23
32	226	18.6	200	2	A26637	neurogenic repetitive	2.40e-23
33	225	18.6	862	2	S43922	versican - pig-tailed	3.69e-23
34	225	18.6	1257	2	S28764	neurocan - rat	3.69e-23
35	225	18.6	2397	2	A55535	versican precursor -	3.69e-23
36	224	18.5	4391	2	A38096	perlecan precursor -	5.66e-23
37	223	18.4	1268	2	S52781	neurocan - mouse	8.69e-23
38	218	18.0	385	2	S53718	homeotic protein dlk	7.37e-22
39	218	18.0	385	2	A54785	preadipocyte factor 1	7.37e-22
40	214	17.7	1295	2	A32901	glp1 protein precurso	4.04e-21
41	207	17.1	473	2	A56175	adhesive plaque prote	7.81e-20
42	207	17.1	1376	2	G00043	osteonidogen - human	7.81e-20
43	205	16.9	1959	1	AGRT	agrin - rat	1.81e-19
44	204	16.8	4544	1	S02392	alpha-2-macroglobulin	2.76e-19
45	201	16.6	260	2	A44549	fetal antigen 1 homeo	9.72e-19

ALIGNMENTS

RESULT 1
ENTRY A31640 #type fragment
TITLE epidermal growth factor-like protein slit - fruit fly (Drosophila melanogaster) (fragment)
ORGANISM #formal_name Drosophila melanogaster
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 14-Aug-1998
ACCESSIONS A31640
REFERENCE A31640
#authors Rothberg, J.M.; Hartley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
#journal Cell (1988) 55:1047-1059
#title slit: An EGF-homologous locus of D. melanogaster involved in the development of the embryonic central nervous system.
#cross-references MUID:89077533
#accession A31640
##molecule_type DNA
##residues 1-530 ##label ROT
##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
#introns 470/3
CLASSIFICATION #superfamily EGF homology
KEYWORDS growth factor
FEATURE
148-181 #domain EGF homology #label EGF
SUMMARY #length 530 #checksum 6330

Query Match 40.5%; Score 491; DB 2; Length 530;
Best Local Similarity 39.5%; Pred. No. 3.64e-76;
Matches 64; Conservative 43; Mismatches 49; Indels 6; Gaps 4;

```
Db   170 CDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMY 229
             |:| :|: | ||::| |||::| ||||: |:| :| | | |:: |:|  ||    ::
Qy     4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db   230 PQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLR 289
           |  |||:  || :|      ||:  :|:| ||: |  || | |::||  :::::: |:
Qy    62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db   290 TRPEANVTI-VFSSGQNGILMYDGQDAHLAVELFNGRIRVSY 330
            :   |:|: ||:: :||||:|:| : |:|| |: |::| ||
Qy   119 NWXRXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160
```

RESULT 2

```
ENTRY           B36665     #type complete
TITLE           slit protein 2 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  16-Dec-1998
ACCESSIONS      B36665
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    B36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1469 ##label ROT
    ##cross-references GB:X53959
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
FEATURE
  66-91             #domain proteoglycan amino-terminal homology #label
                      PAH1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  125-148           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  197-220           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  228-272           #domain proteoglycan carboxyl-terminal homology #label
                      PCS1\
  288-313           #domain proteoglycan amino-terminal homology #label
                      PAH2\
  323-346           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  347-370           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  371-394           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
  395-418           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
  419-442           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
  450-494           #domain proteoglycan carboxyl-terminal homology #label
                      PCS2\
  512-537           #domain proteoglycan amino-terminal homology #label
                      PAH3\
  547-571           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
  572-595           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12\
  596-619           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR13\
  620-643           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR14\
  651-695           #domain proteoglycan carboxyl-terminal homology #label
                      PCS3\
  708-733           #domain proteoglycan amino-terminal homology #label
                      PAH4\
  743-766           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR15\
  767-790           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR16\
  846-890           #domain proteoglycan carboxyl-terminal homology #label
                      PCS4\
  1028-1061         #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match          40.3%;  Score 489;  DB 2;  Length 1469;
  Best Local Similarity  40.5%;  Pred. No. 9.43e-76;
  Matches   66;  Conservative   42;  Mismatches  48;  Indels   7;  Gaps   5;

Db   1050 CDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMY 1109
            |:| :|: | ||::| |||::| ||||: |:| :| | | |:: |:|  ||    ::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db   1110 PQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLR 1169
            |  |||:  || :|       ||:  :|:| ||: |  || | |::||  :::::  |:
Qy     62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db   1170 TRPEANVTI-VFSSAEQNGILMYDGQDAHLAVELFNGRIRVSY 1211
             :   |:|: ||: ||:||||:|:|. : |:|| |: |::| ||
Qy    119 NWXRXNITLQVFT-AEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT    3
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
  66-91             #domain proteoglycan amino-terminal homology #label
                      PAH1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  125-148           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  197-220           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  228-272           #domain proteoglycan carboxyl-terminal homology #label
                      PCS1\
  288-313           #domain proteoglycan amino-terminal homology #label
                      PAH2\
  323-346           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  347-370           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  371-394           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
```

```
   395-418          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   419-442          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   450-494          #domain proteoglycan carboxyl-terminal homology #label
                      PCS2\
   512-537          #domain proteoglycan amino-terminal homology #label
                      PAH3\
   547-571          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   572-595          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12\
   596-619          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR13\
   620-643          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR14\
   651-695          #domain proteoglycan carboxyl-terminal homology #label
                      PCS3\
   708-733          #domain proteoglycan amino-terminal homology #label
                      PAH4\
   743-766          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR15\
   767-790          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR16\
   791-814          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR17\
   815-838          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR18\
   846-890          #domain proteoglycan carboxyl-terminal homology #label
                      PCS4\
  1028-1061         #domain EGF homology #label EGF
SUMMARY          #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           40.3%;  Score 489;  DB 2;  Length 1480;
  Best Local Similarity  40.5%;  Pred. No. 9.43e-76;
  Matches   66;  Conservative   42;  Mismatches  48;  Indels   7;  Gaps   5;

Db   1050 CDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMY 1109
           |:| :|: | ||::| |||::| ||||: |:| :| | | |:: |:|  ||    ::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db   1110 PQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLR 1169
           |  |||:  || :|      ||:  :|:| ||: |  || | |::||  :::::  |:
Qy     62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

     1170 TRPEANVTI-VFSSAEQNGILMYDGQDAHLAVELFNGRIRVSY 1211
           :   |:|: ||: ||:||||:|:| : |:|| |: |::| ||
Qy    119 NWXRXNITLQVFT-AEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT    4
ENTRY           B26637     #type fragment
TITLE           neurogenic repetitive locus 95F protein - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
                  14-Aug-1998
ACCESSIONS      B26637
REFERENCE       A91081
   #authors     Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigel,
                  D.; Vaessin, H.; Campos-Ortega, J.A.
   #journal     EMBO J. (1987) 6:761-766
   #title       EGF homologous sequences encoded in the genome of Drosophila
                  melanogaster, and their relation to neurogenic genes.
   #cross-references MUID:87218537
   #accession   B26637
      ##molecule_type mRNA
      ##residues      1-293 ##label KNU
      ##cross-references GB:X05144; NID:g7519; PID:g929536
GENETICS
   #gene        FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
   216-252           #domain EGF homology #label EGF
SUMMARY          #length 293  #checksum 3413

  Query Match           28.3%;  Score 343;  DB 2;  Length 293;
  Best Local Similarity  42.7%;  Pred. No. 5.00e-46;
  Matches   41;  Conservative   20;  Mismatches  33;  Indels   2;  Gaps   2;

Db    159 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTCEN 218
           |:| ||: |::| :| |:| |: |:||: | | : || | |:| | |  |::  ::   :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    219 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 253
           || :|: |||| |    ||   | |:||| || |:
Qy     64 SPC-EGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    5
ENTRY           A35672     #type complete
TITLE           crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A35672
REFERENCE       A35672
   #authors     Tepass, U.; Theres, C.; Knust, E.
   #journal     Cell (1990) 61:787-799
   #title       crumbs encodes an EGF-like protein expressed on apical
                  membranes of Drosophila epithelial cells and required for
                  organization of epithelia.
   #cross-references MUID:90263104
   #accession   A35672
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2139 ##label TEP
      ##cross-references GB:M33753
      ##note         the authors translated the codon GGC for residue 1928 as
                      Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene        FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
   691-722           #domain EGF homology #label EGF
SUMMARY          #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           28.3%;  Score 343;  DB 2;  Length 2139;
  Best Local Similarity  42.7%;  Pred. No. 5.00e-46;
  Matches   41;  Conservative   20;  Mismatches  33;  Indels   2;  Gaps   2;

Db   1821 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTCEN 1880
           |:| ||: |::| :| |:| |: |:||: | | : || | |:| | |  |::  ::   :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1881 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 1915
           || :|: |||| |    ||   | |:||| || |:
Qy     64 SPC-EGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    6
ENTRY           A40136     #type complete
TITLE           fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
ALTERNATE_NAMES epidermal growth factor homolog precursor
CONTAINS        alternatively spliced fibropellin Ib (EGFI)
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            13-May-1992 #sequence_revision 17-Sep-1997 #text_change
                  07-Aug-1998
ACCESSIONS      A40136; B40136; C40136; A29316; A43131
REFERENCE       A40136
   #authors     Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff,
```

```
                R.A.
   #journal     J. Mol. Evol. (1989) 29:314-327
   #title       Structural analysis of the uEGF gene in the sea urchin
                  Strongylocentrotus purpuratus reveals more similarity to
                  vertebrate than to invertebrate genes with EGF-like
                  repeats.
   #cross-references MUID:90112459
   #accession   A40136
     ##status       preliminary
     ##molecule_type mRNA
     ##residues     1-114 ##label DEL
     ##cross-references GB:X17530; NID:g10225; PID:g667061
   #accession   B40136
     ##status       preliminary; not compared with conceptual translation
     ##molecule_type DNA
     ##residues     181-251,329-370,'R',372-408,'RA',411-441 ##label DE2
   #accession   C40136
     ##status       preliminary; not compared with conceptual translation
     ##molecule_type DNA
     ##residues     'K',747-821,898-978 ##label DE3
REFERENCE       A29316
   #authors     Hursh, D.A.; Andrews, M.E.; Raff, R.A.
   #journal     Science (1987) 237:1487-1490
   #title       A sea urchin gene encodes a polypeptide homologous to
                  epidermal growth factor.
   #cross-references MUID:87319677
   #accession   A29316
     ##status       preliminary
     ##molecule_type mRNA
     ##residues     'S',280-481,786-1064 ##label HUR
     ##cross-references GB:M17421; NID:g161474; PID:g552260
REFERENCE       A43131
   #authors     Hunt, L.T.; Barker, W.C.
   #journal     FASEB J. (1989) 3:1760-1764
   #title       Avidin-like domain in an epidermal growth factor homolog from
                  a sea urchin.
   #cross-references MUID:89196806
   #contents    annotation
COMMENT    EGF homology repeats 10-17 are spliced out in the short form
             (fibropellin Ib).
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-19             #domain signal sequence #status predicted #label SIG\
   20-1064          #product fibropellin I #status predicted #label FIB\
   23-54            #domain EGF homology #label EG01\
   57-175           #domain C1r/C1s repeat homology #label CSR\
   180-211          #domain EGF homology #label EG02\
   218-249          #domain EGF homology #label EG03\
   256-287          #domain EGF homology #label EG04\
   294-325          #domain EGF homology #label EG05\
   332-363          #domain EGF homology #label EG06\
   370-401          #domain EGF homology #label EG07\
   408-439          #domain EGF homology #label EG08\
   446-477          #domain EGF homology #label EG09\
   484-515          #domain EGF homology #label EG10\
   522-553          #domain EGF homology #label EG11\
   560-591          #domain EGF homology #label EG12\
   598-629          #domain EGF homology #label EG13\
   636-667          #domain EGF homology #label EG14\
   674-705          #domain EGF homology #label EG15\
   712-743          #domain EGF homology #label EG16\
   750-781          #domain EGF homology #label EG17\
   788-819          #domain EGF homology #label EG18\
   826-857          #domain EGF homology #label EG19\
   864-895          #domain EGF homology #label EG20\
   902-933          #domain EGF homology #label EG21\
   936-1064         #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439,
   446-457,451-466,
   468-477,484-495    #disulfide_bonds #status predicted\
   489-504,506-515,
   522-533,527-542,
   544-553,560-571,
   565-580,582-591,
   598-609,603-618,
   620-629,636-647,
   641-656,658-667,
   674-685,679-694,
   696-705,712-723,
   717-732,734-743,
   750-761,755-770,
   772-781,788-799,
   793-808,810-819,
   826-837,831-846,
   848-857,864-875,
   869-884,886-895,
   902-913,907-922,
   924-933          #disulfide_bonds #status predicted
SUMMARY         #length 1064  #molecular-weight 112072  #checksum 303

  Query Match          25.4%;  Score 308;  DB 2;  Length 1064;
  Best Local Similarity  46.3%;  Pred. No. 4.27e-39;
  Matches   44;  Conservative   17;  Mismatches  28;  Indels   6;  Gaps   2;

Db    808 CACVPGFTGSNCETNIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCEIS--LDA-- 863
          | |:||::| ||  | |:|    | ||: |:| ||:: | |:  |||  |||:  | |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    864 --CRSMPCQNGATCVNVGADYVCECVPGYAGQNCE 896
            | :  ||||| ||: |   ||:|:||::| :||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    7
ENTRY           I50719     #type complete
TITLE           C-Delta-1 - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  14-Aug-1998
ACCESSIONS      I50719
REFERENCE       I50719
   #authors     Henrique, D.; Adam, J.; Myat, A.; Chitnis, A.; Lewis, J.;
                  Ish-Horowicz, D.
   #journal     Nature (1995) 375:787-790
   #title       Expression of a Delta homologue in prospective neurons in the
                  chick.
   #cross-references MUID:95319507
   #accession   I50719
     ##status       preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type mRNA
     ##residues     1-728 ##label HEN
     ##cross-references EMBL:U26590; NID:g882411; PID:g882412
CLASSIFICATION  #superfamily EGF homology
FEATURE
   454-485          #domain EGF homology #label EGF
SUMMARY         #length 728  #molecular-weight 79861  #checksum 1765

  Query Match          25.1%;  Score 304;  DB 2;  Length 728;
  Best Local Similarity  43.4%;  Pred. No. 2.61e-38;
  Matches   43;  Conservative   21;  Mismatches  29;  Indels   6;  Gaps   4;

Db    436 CQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCPPGYNGKNCSTP--V-S-R 491
          |:| :|::| :| :| |||    | ||: | | || ||| |: ||:|  |  |  : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63
```

```
Db    492 --CEHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
          ||   |:||| | :::|| ||:|  |:|| :|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    8
ENTRY           I48324     #type complete
TITLE           DELTA-like 1 - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                  28-Feb-1997
ACCESSIONS      I48324
REFERENCE       I48324
   #authors     Bettenhausen, B.; de Angelis, M.H.; Simon, D.; Guenet, J.L.;
                  Gossler, A.
   #journal     Development (1995) 121:2407-2418
   #title       Transient and restricted expression during mouse
                  embryogenesis of Dll1, a murine gene closely related to
                  Drosophila Delta.
   #cross-references MUID:95401858
   #accession   I48324
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-722 ##label RES
      ##cross-references EMBL:X80903; NID:g806569; PID:g806570
GENETICS
   #gene        Dll1
SUMMARY         #length 722  #molecular-weight 78448  #checksum 1452

  Query Match           24.9%;  Score 302;  DB 2;  Length 722;
  Best Local Similarity  42.4%;  Pred. No. 6.44e-38;
  Matches    42;  Conservative   22;  Mismatches  29;  Indels   6;  Gaps   5;

Db    428 CRCQAGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGKNCS-AP-V-S-R 483
          | | :|::|  | :| |||    | ||: | | || :|| |: ||:|  |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    484 --CEHAPCHNGATCHQRGQRYMCECAQGYGGPNCQFLLP 520
          || : |:||| | ::| | :|:|  |:|||:|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    9
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
   #authors     Joutel, A.; Tournier-Lasserve, E.
   #submission  submitted to the EMBL Data Library, April 1997
   #accession   S78549
      ##molecule_type mRNA
      ##residues     1-2321 ##label JOU1
      ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE       S71825
   #authors     Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
   #journal     Nature (1996) 383:707-710
   #title       Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
   #cross-references MUID:97032728
   #accession   S71825
      ##status       nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues     67-113;138-194;268-333,'G',335-346;536-613;716-765;
                      1240-1279;1815-1888 ##label JOU2
      ##cross-references EMBL:U97669
GENETICS
   #gene        notch3
   #map_position 19p13.1
FUNCTION
   #description  may be involved in pathogenesis of CADASIL, causing a type of
                   stroke and dementia
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        tandem repeat; transmembrane protein
FEATURE
   318-349           #domain EGF homology #label EGF\
   1838-1870         #domain ankyrin repeat homology #label AN1\
   1871-1903         #domain ankyrin repeat homology #label AN2\
   1905-1937         #domain ankyrin repeat homology #label AN3\
   1938-1970         #domain ankyrin repeat homology #label AN4\
   1971-2003         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2321  #molecular-weight 243657  #checksum 3337

  Query Match           24.6%;  Score 298;  DB 2;  Length 2321;
  Best Local Similarity  42.7%;  Pred. No. 3.91e-37;
  Matches    41;  Conservative   15;  Mismatches  39;  Indels   1;  Gaps   1;

Db   1108 CECLPGYNGDNCEDDVDECASQPCQHGGSCIDLVARYLCSCPPGTLGVLCEINEDDCGPG 1167
          |||:||| |||| :: |:|  : ||:|: |:| |  | | |: |  | ||||     :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1168 PPLDSGPRCLHNGTCVDLVGGFRCTCPPGYTGLRCE 1203
          :| : |  | : : |||      | | ||: |  ||
Qy     64 SPCE-GTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   10
ENTRY           A49175     #type fragment
TITLE           Motch B protein - mouse (fragment)
ALTERNATE_NAMES  Notch homolog
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                  14-Aug-1998
ACCESSIONS      A49175; PH1570; S32113
REFERENCE       A49175
   #authors     Lardelli, M.; Lendahl, U.
   #journal     Exp. Cell Res. (1993) 204:364-372
   #title       Motch A and Motch B--two mouse Notch homologues coexpressed
                  in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession   A49175
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues     1-1203 ##label LAR
      ##cross-references EMBL:X68279; NID:g287989; PID:g287990
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:126158)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermaland neural fate for cells in the early
             embryo.
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   560-591           #domain EGF homology #label EGF
SUMMARY         #length 1203  #checksum 910

  Query Match           24.3%;  Score 294;  DB 2;  Length 1203;
  Best Local Similarity  44.8%;  Pred. No. 2.36e-36;
  Matches    43;  Conservative   17;  Mismatches  31;  Indels   5;  Gaps   4;

Db    855 RCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCE--ENI-D- 910
          ||||:||| | ||  : |:|::: ||||: |:| || : | |: |  | |||   ::
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    911 ECA-GGPHCLNGGQCVDRIGGYTCRCLPGFAGERCE 945
          :  |  | ||: |||:     |:|||||:|  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98
```

```
RESULT  11
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
   #authors     Bierkamp, C.; Campos-Ortega, J.A.
   #journal     Mech. Dev. (1993) 43:87-100
   #title       A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
   #cross-references MUID:94128602
   #accession   S42612
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2437 ##label BIE
      ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1915-1947          #domain ankyrin repeat homology #label AN1\
   1948-1980          #domain ankyrin repeat homology #label AN2\
   1982-2014          #domain ankyrin repeat homology #label AN3\
   2015-2047          #domain ankyrin repeat homology #label AN4\
   2048-2080          #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           24.0%;  Score 291;  DB 2;  Length 2437;
  Best Local Similarity  44.8%;  Pred. No. 9.10e-36;
  Matches   43;  Conservative   15;  Mismatches  31;  Indels   7;  Gaps   5;

Db    474 HCICMPGYEGVFCQINSDDCASQPCLNG-KCIDKINSFHCECPKGFSGSLCQV--DV-D- 528
          :| ||||| |  |  | |||  : | ||  |:| :||: | |: |:|| ||::   :
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    529 -E-CASTPCKNGAKCTDGPNKYTCECTPGFSGIHCE 562
           | :| | |||:| |  |:  |:| |||:|  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT  12
ENTRY           A48836     #type complete
TITLE           fibropellin C precursor - sea urchin (Strongylocentrotus
                  purpuratus)
ALTERNATE_NAMES EGF repeat-containing protein; epidermal growth
                  factor-related protein 3; fibropellin III
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  07-Aug-1998
ACCESSIONS      A48836
REFERENCE       A48836
   #authors     Bisgrove, B.W.; Raff, R.A.
   #journal     Dev. Biol. (1993) 157:526-538
   #title       The SpEGF III gene encodes a member of the fibropellins: EGF
                  repeat-containing proteins that form the apical lamina of
                  the sea urchin embryo.
   #cross-references MUID:93273088
   #accession   A48836
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-570 ##label BIS
      ##cross-references GB:L07045; NID:g310659; PID:g310660
      ##note         sequence extracted from NCBI backbone (NCBIN:132724,
                      NCBIP:132725)
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-18               #domain signal sequence #status predicted #label SIG\
   19-570             #product fibropellin C #status predicted #label FIB\
   19-54              #domain EGF homology #label EGF1\
   57-175             #domain C1r/C1s repeat homology #label C1R\
   176-211            #domain EGF homology #label EGF2\
   214-249            #domain EGF homology #label EGF3\
   252-287            #domain EGF homology #label EGF4\
   290-325            #domain EGF homology #label EGF5\
   328-363            #domain EGF homology #label EGF6\
   366-401            #domain EGF homology #label EGF7\
   404-439            #domain EGF homology #label EGF8\
   442-570            #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439    #disulfide_bonds #status predicted
SUMMARY         #length 570  #molecular-weight 61115  #checksum 5567

  Query Match           23.7%;  Score 287;  DB 2;  Length 570;
  Best Local Similarity  42.1%;  Pred. No. 5.47e-35;
  Matches   40;  Conservative   18;  Mismatches  31;  Indels   6;  Gaps   2;

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYVCQCLPNYTGTHCEIS--LDA-- 369
          |:| :|::| ||  | ::|    | ||: |:| |::| | |   |:|  |||:  | |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    370 --CASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402
            | :  ||||: | : |   ||:||||: | :||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT  13
ENTRY           A49128     #type complete
TITLE           cell-fate determining gene Notch2 protein - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A49128
REFERENCE       A49128
   #authors     Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal     Development (1992) 116:931-941
   #title.      Notch2: a second mammalian Notch gene.
   #cross-references MUID:93202015
   #accession   A49128
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-2471 ##label WEI
      ##experimental_source Schwann cell
      ##note         sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   1029-1060          #domain EGF homology #label EGF\
   1876-1908          #domain ankyrin repeat homology #label AN1\
   1909-1941          #domain ankyrin repeat homology #label AN2\
   1943-1975          #domain ankyrin repeat homology #label AN3\
   1976-2008          #domain ankyrin repeat homology #label AN4\
   2009-2041          #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           23.7%;  Score 287;  DB 2;  Length 2471;
  Best Local Similarity  45.4%;  Pred. No. 5.47e-35;
  Matches   44;  Conservative   17;  Mismatches  29;  Indels   7;  Gaps   5;

Db   1172 RCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCE--ENI-D- 1227
          ||||:||| | ||  : |:|::: ||||: |:| || : | |: |  | |||   ::
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62
```

```
Db   1228 -D-CAGAPHCLNGGQCVDRIGGYSCRCLPGFAGERCE 1262
           | |:  | ||: |||:     |:|||||:|  ||
Qy     63 KSPCEGT-ECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   14
ENTRY            A56136     #type complete
TITLE            jagged protein precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change
                   11-Aug-1995
ACCESSIONS       A56136
REFERENCE        A56136
   #authors      Lindsell, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.
   #journal      Cell (1995) 80:909-917
   #title        Jagged: a mammalian ligand that activates Notch1.
   #cross-references MUID:95211842
   #accession    A56136
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-1220 ##label LIN
      ##cross-references GB:L38483
SUMMARY          #length 1220  #molecular-weight 134528  #checksum 2746

  Query Match           23.3%;  Score 283;  DB 2;  Length 1220;
  Best Local Similarity  40.8%;  Pred. No. 3.28e-34;
  Matches    40;  Conservative   18;  Mismatches  34;  Indels   6;  Gaps   3;

Db    473 RCICPPGYAGDHCERDIDECASNPCLNGGHCQNEINRFQCLCPTGFSGNLC----QLDI- 527
          || | ||||||:|  : |:|  : | ||::| :|:| : |||: |:|| ||    :|
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    528 DY-CEPNPCQNGAQCYNRASDYFCKCPEDYEGKNCSHL 564
           ||   ||||| | ::::   | |   : | :|  |
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKL 100


RESULT   15
ENTRY            S45306     #type complete
TITLE            notch 3 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   10-Jul-1998
ACCESSIONS       S45306
REFERENCE        S45306
   #authors      Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal      Mech. Dev. (1994) 46:123-136
   #title        The novel Notch homologue mouse Notch 3 lacks specific
                   epidermal growth factor-repeats and is expressed in
                   proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession    S45306
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1839-1871          #domain ankyrin repeat homology #label AN1\
   1872-1904          #domain ankyrin repeat homology #label AN2\
   1906-1938          #domain ankyrin repeat homology #label AN3\
   1939-1971          #domain ankyrin repeat homology #label AN4\
   1972-2004          #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           23.0%;  Score 279;  DB 2;  Length 2318;
  Best Local Similarity  41.7%;  Pred. No. 1.95e-33;
  Matches    40;  Conservative   20;  Mismatches  30;  Indels   6;  Gaps   2;

Db    456 CICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMC----QLDV-- 509
          | ||:|::|  |  : |:|:   | ||: |:| ||::|| |: |:|| :|    :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    510 DECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCER 545
           | :| |:|||:|||| :   |:|  || |  ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEK 99


Search completed: Fri May 28 08:54:12 1999
Job time : 29 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:54:30 1999; MasPar time 6.51 Seconds
694.790 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-5
Description: (1-160) from US09191647.pep
Perfect Score: 1212
Sequence: 1 WPRCECMPGYAGDNCSENQD.....NGGNDHIAVXLYXGHVRFSY 160

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 41.676; Variance 70.700; scale 0.589

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	489	40.3	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	3.30e-86
2	343	28.3	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	1.92e-52
3	308	25.4	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	1.40e-44
4	304	25.1	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	1.09e-43
5	302	24.9	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	3.04e-43
6	295	24.3	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	1.09e-41
7	291	24.0	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	8.36e-41
8	287	23.7	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	6.39e-40
9	279	23.0	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	3.68e-38
10	273	22.5	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	7.61e-37
11	272	22.4	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	1.26e-36
12	272	22.4	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	1.26e-36
13	269	22.2	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	5.69e-36
14	269	22.2	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	5.69e-36
15	268	22.1	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	9.41e-36
16	246	20.3	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	5.44e-31
17	241	19.9	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	6.42e-30
18	231	19.1	3562	1	PGCV_CHICK	VERSICAN CORE PROTEIN	8.66e-28
19	229	18.9	3396	1	PGCV_HUMAN	VERSICAN CORE PROTEIN	2.30e-27
20	227	18.7	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	6.08e-27
21	225	18.6	862	1	PGCV_MACNE	VERSICAN CORE PROTEIN	1.61e-26
22	225	18.6	1257	1	PGCN_RAT	NEUROCAN CORE PROTEIN	1.61e-26
23	225	18.6	3358	1	PGCV_MOUSE	VERSICAN CORE PROTEIN	1.61e-26
24	224	18.5	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	2.61e-26
25	223	18.4	1268	1	PGCN_MOUSE	NEUROCAN CORE PROTEIN	4.24e-26
26	218	18.0	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	4.76e-25
27	214	17.7	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	3.27e-24
28	207	17.1	1376	1	NID2_HUMAN	NIDOGEN-2 PRECURSOR (N	9.31e-23
29	205	16.9	1959	1	AGRI_RAT	AGRIN PRECURSOR.	2.41e-22
30	204	16.8	4544	1	LRP1_HUMAN	LOW-DENSITY LIPOPROTEI	3.88e-22
31	201	16.6	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	1.61e-21
32	198	16.3	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	6.65e-21
33	193	15.9	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	6.98e-20
34	188	15.5	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	7.21e-19
35	181	14.9	432	1	NEL2_RAT	NEL-LIKE PROTEIN (FRAG	1.85e-17
36	179	14.8	2109	1	PGCA_CHICK	AGGRECAN CORE PROTEIN	4.63e-17
37	178	14.7	515	1	APX1_CAEEL	APX-1 PROTEIN PRECURSO	7.33e-17
38	172	14.2	883	1	PGCB_RAT	BREVICAN CORE PROTEIN	1.13e-15
39	171	14.1	564	1	PGCA_CANFA	AGGRECAN CORE PROTEIN	1.79e-15
40	171	14.1	883	1	PGCB_MOUSE	BREVICAN CORE PROTEIN	1.79e-15
41	165	13.6	4543	1	LRP1_CHICK	LOW-DENSITY LIPOPROTEI	2.67e-14
42	164	13.5	886	1	EMR1_HUMAN	CELL SURFACE GLYCOPROT	4.18e-14
43	161	13.3	428	1	NEL2_HUMAN	NEL-LIKE PROTEIN (FRAG	1.60e-13
44	161	13.3	2871	1	FBN1_BOVIN	FIBRILLIN 1 PRECURSOR	1.60e-13
45	160	13.2	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	2.49e-13

ALIGNMENTS

```
RESULT    1
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE  1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          40.3%;  Score 489;  DB 1;  Length 1480;
  Best Local Similarity  40.5%;  Pred. No. 3.30e-86;
  Matches   66;  Conservative   42;  Mismatches  48;  Indels   7;  Gaps   5;

Db   1050 CDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMY 1109
            |:| :|: | ||::| |||::| |||: |:| :| | | |:: |:|  ||    ::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db   1110 PQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLR 1169
             | |||: .|| :|       ||:  :|:| ||: |  || | |::||  :::::  |:
Qy     62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db   1170 TRPEANVTI-VFSSAEQNGILMYDGQDAHLAVELFNGRIRVSY 1211
                :   |:|: ||: ||:|||||:|:| : |:|| |: |::| ||
Qy    119 NWXRXNITLQVFT-AEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT   2
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
```

```
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1    90
FT   CHAIN       91  2139       CRUMBS PROTEIN.
FT   DOMAIN      91  2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085  2111       POTENTIAL.
FT   DOMAIN    2112  2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267   303       EGF-LIKE 1.
FT   DOMAIN     306   343       EGF-LIKE 2.
FT   DOMAIN     348   386       EGF-LIKE 3.
FT   DOMAIN     388   425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427   463       EGF-LIKE 5.
FT   DOMAIN     464   500       EGF-LIKE 6.
FT   DOMAIN     501   532       EGF-LIKE 7.
FT   DOMAIN     545   581       EGF-LIKE 8.
FT   DOMAIN     582   611       EGF-LIKE 9.
FT   DOMAIN     609   646       EGF-LIKE 10.
FT   DOMAIN     648   685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687   723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725   761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763   800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802   838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840   902       EGF-LIKE 16.
FT   DOMAIN     904   940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942   978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980  1021       EGF-LIKE 19.
FT   DOMAIN    1207  1243       EGF-LIKE 20.
FT   DOMAIN    1481  1517       EGF-LIKE 21.
FT   DOMAIN    1759  1795       EGF-LIKE 22.
FT   DOMAIN    1797  1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835  1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874  1915       EGF-LIKE 25.
FT   DOMAIN    1915  1951       EGF-LIKE 26.
FT   DOMAIN    1953  1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991  2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030  2070       EGF-LIKE 29.
FT   DISULFID   271   282       BY SIMILARITY.
FT   DISULFID   276   291       BY SIMILARITY.
FT   DISULFID   293   302       BY SIMILARITY.
FT   DISULFID   310   321       BY SIMILARITY.
FT   DISULFID   315   331       BY SIMILARITY.
FT   DISULFID   333   342       BY SIMILARITY.
FT   DISULFID   352   363       BY SIMILARITY.
FT   DISULFID   357   374       BY SIMILARITY.
FT   DISULFID   376   385       BY SIMILARITY.
FT   DISULFID   392   403       BY SIMILARITY.
FT   DISULFID   397   412       BY SIMILARITY.
FT   DISULFID   414   424       BY SIMILARITY.
FT   DISULFID   431   442       BY SIMILARITY.
FT   DISULFID   436   451       BY SIMILARITY.
FT   DISULFID   453   462       BY SIMILARITY.
FT   DISULFID   468   479       BY SIMILARITY.
FT   DISULFID   473   488       BY SIMILARITY.
FT   DISULFID   490   499       BY SIMILARITY.
FT   DISULFID   505   515       BY SIMILARITY.
FT   DISULFID   509   520       BY SIMILARITY.
FT   DISULFID   522   531       BY SIMILARITY.
FT   DISULFID   549   562       BY SIMILARITY.
FT   DISULFID   556   569       BY SIMILARITY.
FT   DISULFID   571   580       BY SIMILARITY.
FT   DISULFID   586   597       BY SIMILARITY.
FT   DISULFID   591   602       BY SIMILARITY.
FT   DISULFID   604   610       BY SIMILARITY.
FT   DISULFID   613   624       BY SIMILARITY.
FT   DISULFID   618   634       BY SIMILARITY.
FT   DISULFID   636   645       BY SIMILARITY.
FT   DISULFID   652   664       BY SIMILARITY.
FT   DISULFID   659   673       BY SIMILARITY.
FT   DISULFID   675   684       BY SIMILARITY.
FT   DISULFID   691   702       BY SIMILARITY.
FT   DISULFID   696   711       BY SIMILARITY.
FT   DISULFID   713   722       BY SIMILARITY.
FT   DISULFID   729   740       BY SIMILARITY.
FT   DISULFID   734   749       BY SIMILARITY.
FT   DISULFID   751   760       BY SIMILARITY.
FT   DISULFID   767   778       BY SIMILARITY.
FT   DISULFID   772   787       BY SIMILARITY.
FT   DISULFID   789   799       BY SIMILARITY.
FT   DISULFID   806   817       BY SIMILARITY.
FT   DISULFID   811   826       BY SIMILARITY.
FT   DISULFID   828   837       BY SIMILARITY.
FT   DISULFID   844   855       BY SIMILARITY.
FT   DISULFID   849   890       BY SIMILARITY.
FT   DISULFID   892   901       BY SIMILARITY.
FT   DISULFID   908   919       BY SIMILARITY.
FT   DISULFID   913   928       BY SIMILARITY.
FT   DISULFID   930   939       BY SIMILARITY.
FT   DISULFID   946   957       BY SIMILARITY.
FT   DISULFID   952   966       BY SIMILARITY.
FT   DISULFID   968   977       BY SIMILARITY.
FT   DISULFID   984   995       BY SIMILARITY.
FT   DISULFID   989  1009       BY SIMILARITY.
FT   DISULFID  1011  1020       BY SIMILARITY.
FT   DISULFID  1211  1222       BY SIMILARITY.
FT   DISULFID  1216  1231       BY SIMILARITY.
FT   DISULFID  1233  1242       BY SIMILARITY.
FT   DISULFID  1485  1496       BY SIMILARITY.
FT   DISULFID  1490  1505       BY SIMILARITY.
FT   DISULFID  1507  1516       BY SIMILARITY.
FT   DISULFID  1763  1774       BY SIMILARITY.
FT   DISULFID  1768  1783       BY SIMILARITY.
FT   DISULFID  1785  1794       BY SIMILARITY.
FT   DISULFID  1801  1812       BY SIMILARITY.
FT   DISULFID  1806  1821       BY SIMILARITY.
FT   DISULFID  1823  1832       BY SIMILARITY.
FT   DISULFID  1839  1850       BY SIMILARITY.
FT   DISULFID  1844  1859       BY SIMILARITY.
FT   DISULFID  1861  1870       BY SIMILARITY.
FT   DISULFID  1878  1889       BY SIMILARITY.
FT   DISULFID  1883  1903       BY SIMILARITY.
FT   DISULFID  1905  1914       BY SIMILARITY.
FT   DISULFID  1919  1930       BY SIMILARITY.
FT   DISULFID  1924  1939       BY SIMILARITY.
FT   DISULFID  1941  1950       BY SIMILARITY.
FT   DISULFID  1957  1968       BY SIMILARITY.
FT   DISULFID  1962  1977       BY SIMILARITY.
FT   DISULFID  1979  1988       BY SIMILARITY.
FT   DISULFID  1995  2008       BY SIMILARITY.
FT   DISULFID  2002  2017       BY SIMILARITY.
FT   DISULFID  2019  2028       BY SIMILARITY.
FT   CARBOHYD    37    37       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
```

```
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    238    238       POTENTIAL.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    336    336       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CARBOHYD    550    550       POTENTIAL.
FT   CARBOHYD    565    565       POTENTIAL.
FT   CARBOHYD    736    736       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    860    860       POTENTIAL.
FT   CARBOHYD    884    884       POTENTIAL.
FT   CARBOHYD    976    976       POTENTIAL.
FT   CARBOHYD   1102   1102       POTENTIAL.
FT   CARBOHYD   1114   1114       POTENTIAL.
FT   CARBOHYD   1138   1138       POTENTIAL.
FT   CARBOHYD   1192   1192       POTENTIAL.
FT   CARBOHYD   1245   1245       POTENTIAL.
FT   CARBOHYD   1255   1255       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   CARBOHYD   1363   1363       POTENTIAL.
FT   CARBOHYD   1441   1441       POTENTIAL.
FT   CARBOHYD   1454   1454       POTENTIAL.

Note: remainder of annotations omitted.

  Query Match           28.3%;  Score 343;  DB 1;  Length 2139;
  Best Local Similarity  42.7%;  Pred. No. 1.92e-52;
  Matches   41;  Conservative   20;  Mismatches  33;  Indels   2;  Gaps   2;

Db   1821 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPPEDYMGPQCDVLKQMTCEN 1880
            |:| ||: |::| :| |:| |: |:||: | | : || | |:| | |  |::  ::   :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1881 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 1915
             || :|: |||| |    ||   | |:||| || |:
Qy     64 SPC-EGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    3
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1064       FIBROPELLIN I.
FT   DOMAIN      20     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480    516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     518    554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     556    592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     594    630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632    668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     670    706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     708    744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     746    782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     784    820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     822    858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      860    896       EGF-LIKE 20.
FT   DOMAIN      898    934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      936   1064       AVIDIN-LIKE.
FT   DISULFID     23     34       BY SIMILARITY.
FT   DISULFID     28     43       BY SIMILARITY.
FT   DISULFID     45     54       BY SIMILARITY.
FT   DISULFID    180    191       BY SIMILARITY.
FT   DISULFID    185    200       BY SIMILARITY.
FT   DISULFID    202    211       BY SIMILARITY.
FT   DISULFID    218    229       BY SIMILARITY.
FT   DISULFID    223    238       BY SIMILARITY.
FT   DISULFID    240    249       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    261    276       BY SIMILARITY.
FT   DISULFID    278    287       BY SIMILARITY.
FT   DISULFID    294    305       BY SIMILARITY.
FT   DISULFID    299    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   DISULFID    446    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    477       BY SIMILARITY.
FT   DISULFID    484    495       BY SIMILARITY.
FT   DISULFID    489    504       BY SIMILARITY.
FT   DISULFID    506    515       BY SIMILARITY.
FT   DISULFID    522    533       BY SIMILARITY.
FT   DISULFID    527    542       BY SIMILARITY.
FT   DISULFID    544    553       BY SIMILARITY.
FT   DISULFID    560    571       BY SIMILARITY.
FT   DISULFID    565    580       BY SIMILARITY.
FT   DISULFID    582    591       BY SIMILARITY.
FT   DISULFID    598    609       BY SIMILARITY.
FT   DISULFID    603    618       BY SIMILARITY.
FT   DISULFID    620    629       BY SIMILARITY.
FT   DISULFID    636    647       BY SIMILARITY.
FT   DISULFID    641    656       BY SIMILARITY.
FT   DISULFID    658    667       BY SIMILARITY.
FT   DISULFID    674    685       BY SIMILARITY.
FT   DISULFID    679    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    712    723       BY SIMILARITY.
FT   DISULFID    717    732       BY SIMILARITY.
FT   DISULFID    734    743       BY SIMILARITY.
FT   DISULFID    750    761       BY SIMILARITY.
FT   DISULFID    755    770       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    788    799       BY SIMILARITY.
FT   DISULFID    793    808       BY SIMILARITY.
FT   DISULFID    810    819       BY SIMILARITY.
FT   DISULFID    826    837       BY SIMILARITY.
FT   DISULFID    831    846       BY SIMILARITY.
FT   DISULFID    848    857       BY SIMILARITY.
FT   DISULFID    864    875       BY SIMILARITY.
FT   DISULFID    869    884       BY SIMILARITY.
FT   DISULFID    886    895       BY SIMILARITY.
FT   DISULFID    902    913       BY SIMILARITY.
FT   DISULFID    907    922       BY SIMILARITY.
FT   DISULFID    924    933       BY SIMILARITY.
FT   VARSPLIC    477    780       MISSING (IN FORM IB).
FT   CARBOHYD     30     30       POTENTIAL.
FT   CARBOHYD    136    136       POTENTIAL.
FT   CARBOHYD    851    851       POTENTIAL.
FT   CONFLICT    279    279       L -> S (IN REF. 2).
SQ   SEQUENCE   1064 AA;  112072 MW;  FBD10D48 CRC32;
```

```
  Query Match          25.4%;  Score 308;  DB 1;  Length 1064;
  Best Local Similarity  46.3%;  Pred. No. 1.40e-44;
  Matches   44;  Conservative   17;  Mismatches  28;  Indels   6;  Gaps   2;

Db    808 CACVPGFTGSNCETNIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCEIS--LDA-- 863
          | |:||::| ||  | |:|    | ||: |:| ||:: | |:  |||  |||:  | |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    864 --CRSMPCQNGATCVNVGADYVCECVPGYAGQNCE 896
            | :  ||||| ||: |   ||:|:||::| :||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    4
ID   DLL1_HUMAN     STANDARD;      PRT;   723 AA.
AC   O00548;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MANN R.S., GRAY G.E., HENRIQUE D., ISH-HOROWICZ D.,
RA   ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF003522; G2197069; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    723       DELTA-LIKE PROTEIN 1.
FT   DOMAIN       18    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    568       POTENTIAL.
FT   DOMAIN      569    723       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      226    254       EGF-LIKE 1.
FT   DOMAIN      257    285       EGF-LIKE 2.
FT   DOMAIN      292    325       EGF-LIKE 3.
FT   DOMAIN      332    363       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN      370    402       EGF-LIKE 5.
FT   DOMAIN      409    440       EGF-LIKE 6.
FT   DOMAIN      447    478       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      485    516       EGF-LIKE 8.
FT   DISULFID    226    237       BY SIMILARITY.
FT   DISULFID    230    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    257    268       BY SIMILARITY.
FT   DISULFID    263    274       BY SIMILARITY.
FT   DISULFID    276    285       BY SIMILARITY.
FT   DISULFID    292    304       BY SIMILARITY.
```

```
FT   DISULFID    298    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    391       BY SIMILARITY.
FT   DISULFID    393    402       BY SIMILARITY.
FT   DISULFID    409    420       BY SIMILARITY.
FT   DISULFID    414    429       BY SIMILARITY.
FT   DISULFID    431    440       BY SIMILARITY.
FT   DISULFID    447    467       BY SIMILARITY.
FT   DISULFID    469    478       BY SIMILARITY.
FT   DISULFID    485    496       BY SIMILARITY.
FT   DISULFID    490    505       BY SIMILARITY.
FT   DISULFID    507    516       BY SIMILARITY.
FT   CARBOHYD    477    477       POTENTIAL.
SQ   SEQUENCE   723 AA;  77956 MW;  A1D48BDB CRC32;

  Query Match           25.1%;  Score 304;  DB 1;  Length 723;
  Best Local Similarity  42.4%;  Pred. No. 1.09e-43;
  Matches   42;  Conservative   24;  Mismatches  27;  Indels   6;  Gaps   5;

Db    429 CRCQAGFSGRHCDDNVDDCASSPCANGGTCRDGVNDFSCTCPPGYTGRNCS-AP-V-S-R 484
          | | :|::| :| :| |||    | ||: | | || :|| |: ||:|: |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    485 --CEHAPCHNGATCHERGHGYVCECARGYGGPNCQFLLP 521
           || : |:||| | ::|:  ||:|  |:|||:|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    5
ID   DLL1_MOUSE     STANDARD;      PRT;   722 AA.
AC   Q61483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C X C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95401858.
RA   BETTENHAUSEN B., DE ANGELIS M.H., SIMON D., GUENET J.-L., GOSSLER A.;
RT   "Transient and restricted expression during mouse embryogenesis of
RT   Dll1, a murine gene closely related to Drosophila Delta.";
RT   DEVELOPMENT 121:2407-2418(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN THE PARAXIAL
CC       MESODERM AND NERVOUS SYSTEM. EXPRESSED AT HIGH LEVELS IN ADULT
CC       HEART AND AT LOWER LEVELS, IN ADULT LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED UNTIL DAY 15 IN THE EMBRYO.
CC       EXPRESSION THEN DECREASES AND INCREASES AGAIN IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80903; G806570; -.
DR   MGD; MGI:104659; DLL1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    722       DELTA-LIKE PROTEIN 1.
FT   DOMAIN       18    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    568       POTENTIAL.
FT   DOMAIN      569    722       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      225    253       EGF-LIKE 1.
FT   DOMAIN      256    284       EGF-LIKE 2.
FT   DOMAIN      291    324       EGF-LIKE 3.
FT   DOMAIN      331    362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN      369    401       EGF-LIKE 5.
FT   DOMAIN      408    439       EGF-LIKE 6.
FT   DOMAIN      446    477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      484    515       EGF-LIKE 8.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    229    242       BY SIMILARITY.
FT   DISULFID    244    253       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    262    273       BY SIMILARITY.
FT   DISULFID    275    284       BY SIMILARITY.
FT   DISULFID    291    303       BY SIMILARITY.
FT   DISULFID    297    313       BY SIMILARITY.
FT   DISULFID    315    324       BY SIMILARITY.
FT   DISULFID    331    342       BY SIMILARITY.
FT   DISULFID    336    351       BY SIMILARITY.
FT   DISULFID    353    362       BY SIMILARITY.
FT   DISULFID    369    380       BY SIMILARITY.
FT   DISULFID    374    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   DISULFID    446    466       BY SIMILARITY.
FT   DISULFID    468    477       BY SIMILARITY.
FT   DISULFID    484    495       BY SIMILARITY.
FT   DISULFID    489    504       BY SIMILARITY.
FT   DISULFID    506    515       BY SIMILARITY.
FT   CARBOHYD    476    476       POTENTIAL.
SQ   SEQUENCE   722 AA;  78448 MW;  5A647702 CRC32;

  Query Match           24.9%;  Score 302;  DB 1;  Length 722;
  Best Local Similarity  42.4%;  Pred. No. 3.04e-43;
  Matches   42;  Conservative   22;  Mismatches  29;  Indels   6;  Gaps   5;

Db    428 CRCQAGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGKNCS-AP-V-S-R 483
          | | :|::|  | :| |||    | ||: | | || :|| |: ||:|  |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    484 --CEHAPCHNGATCHQRGQRYMCECAQGYGGPNCQFLLP 520
           || : |:||| | ::| | :|:|  |:|||:|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    6
ID   DLL1_RAT       STANDARD;      PRT;   714 AA.
AC   P97677;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DISIBIO G., HEBSHI L., BOULTER J., WEINMASTER G.;
```

```
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78889; G1699046; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   714       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   537       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   538   560       POTENTIAL.
FT   DOMAIN     561   714       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225   253       EGF-LIKE 1.
FT   DOMAIN     256   284       EGF-LIKE 2.
FT   DOMAIN     291   324       EGF-LIKE 3.
FT   DOMAIN     331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369   401       EGF-LIKE 5.
FT   DOMAIN     408   439       EGF-LIKE 6.
FT   DOMAIN     446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484   515       EGF-LIKE 8.
FT   DISULFID   225   236       BY SIMILARITY.
FT   DISULFID   229   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   262   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   303       BY SIMILARITY.
FT   DISULFID   297   313       BY SIMILARITY.
FT   DISULFID   315   324       BY SIMILARITY.
FT   DISULFID   331   342       BY SIMILARITY.
FT   DISULFID   336   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   380       BY SIMILARITY.
FT   DISULFID   374   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   CARBOHYD   476   476       POTENTIAL.
SQ   SEQUENCE   714 AA;  77378 MW;  604B76D1 CRC32;

  Query Match          24.3%;  Score 295;  DB 1;  Length 714;
  Best Local Similarity  41.4%;  Pred. No. 1.09e-41;
  Matches   41;  Conservative   23;  Mismatches  29;  Indels   6;  Gaps   5;

Db    428 CRCQTGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGRNCS-AP-V-S-R 483
          | |  |::|  | :| |||    | ||: | | || :|| |: ||:|: |  :| : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    484 --CEHAPCHNGATCHQRGQRYMCECAQGYGGANCQFLLP 520
```

```
                || : |:||| | ::| | :|:|  |:||::|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT   7
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    20       POTENTIAL.
FT   CHAIN       21  2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21  1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725  1747       POTENTIAL.
FT   DOMAIN    1748  2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21    57       EGF-LIKE 1.
FT   DOMAIN      58    98       EGF-LIKE 2.
FT   DOMAIN     101   138       EGF-LIKE 3.
FT   DOMAIN     139   175       EGF-LIKE 4.
FT   DOMAIN     177   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217   254       EGF-LIKE 6.
FT   DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      371    409       EGF-LIKE 10.
FT   DOMAIN      411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      827    865       EGF-LIKE 22.
FT   DOMAIN      867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      981   1017       EGF-LIKE 26.
FT   DOMAIN     1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1057   1093       EGF-LIKE 28.
FT   DOMAIN     1095   1141       EGF-LIKE 29.
FT   DOMAIN     1143   1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1181   1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1219   1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1265   1303       EGF-LIKE 33.
FT   DOMAIN     1305   1344       EGF-LIKE 34.
FT   DOMAIN     1346   1382       EGF-LIKE 35.
FT   DOMAIN     1385   1423       EGF-LIKE 36.
FT   DOMAIN     1446   1561       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1486       LIN/NOTCH 1.
FT   REPEAT     1487   1520       LIN/NOTCH 2.
FT   REPEAT     1521   1561       LIN/NOTCH 3.
FT   DOMAIN     1861   2074       6 X ANK MOTIF REPEATS.
FT   REPEAT     1861   1891       ANK MOTIF 1.
FT   REPEAT     1892   1940       ANK MOTIF 1.
FT   REPEAT     1941   1974       ANK MOTIF 1.
FT   REPEAT     1975   2007       ANK MOTIF 1.
FT   REPEAT     2008   2040       ANK MOTIF 1.
FT   REPEAT     2041   2074       ANK MOTIF 1.
FT   DOMAIN     2265   2276       POLY-GLN (OPA-REPEAT).
FT   DISULFID     25     35       BY SIMILARITY.
FT   DISULFID     29     45       BY SIMILARITY.
FT   DISULFID     47     56       BY SIMILARITY.
FT   DISULFID     62     73       BY SIMILARITY.
FT   DISULFID     67     86       BY SIMILARITY.
FT   DISULFID     88     97       BY SIMILARITY.
FT   DISULFID    105    116       BY SIMILARITY.
FT   DISULFID    110    126       BY SIMILARITY.
FT   DISULFID    128    137       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    163       BY SIMILARITY.
FT   DISULFID    165    174       BY SIMILARITY.
FT   DISULFID    181    194       BY SIMILARITY.
FT   DISULFID    188    203       BY SIMILARITY.
FT   DISULFID    205    214       BY SIMILARITY.
FT   DISULFID    221    232       BY SIMILARITY.
FT   DISULFID    226    242       BY SIMILARITY.
FT   DISULFID    244    253       BY SIMILARITY.
FT   DISULFID    260    271       BY SIMILARITY.
FT   DISULFID    265    280       BY SIMILARITY.
FT   DISULFID    282    291       BY SIMILARITY.
FT   DISULFID    298    311       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    322    331       BY SIMILARITY.
FT   DISULFID    338    349       BY SIMILARITY.
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    375    386       BY SIMILARITY.
FT   DISULFID    380    397       BY SIMILARITY.
FT   DISULFID    399    408       BY SIMILARITY.
FT   DISULFID    415    428       BY SIMILARITY.
FT   DISULFID    422    437       BY SIMILARITY.
FT   DISULFID    439    448       BY SIMILARITY.
FT   DISULFID    455    466       BY SIMILARITY.
FT   DISULFID    460    475       BY SIMILARITY.
FT   DISULFID    477    486       BY SIMILARITY.
FT   DISULFID    493    503       BY SIMILARITY.
FT   DISULFID    498    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    578       BY SIMILARITY.
FT   DISULFID    573    587       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    605    616       BY SIMILARITY.
FT   DISULFID    610    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    653       BY SIMILARITY.
FT   DISULFID    648    662       BY SIMILARITY.
FT   DISULFID    664    673       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   DISULFID    685    700       BY SIMILARITY.
FT   DISULFID    702    711       BY SIMILARITY.
FT   DISULFID    718    728       BY SIMILARITY.
FT   DISULFID    723    737       BY SIMILARITY.
FT   DISULFID    739    748       BY SIMILARITY.
FT   DISULFID    755    766       BY SIMILARITY.
FT   DISULFID    760    775       BY SIMILARITY.
FT   DISULFID    777    786       BY SIMILARITY.
FT   DISULFID    793    804       BY SIMILARITY.
FT   DISULFID    798    813       BY SIMILARITY.
FT   DISULFID    815    824       BY SIMILARITY.
FT   DISULFID    831    842       BY SIMILARITY.
FT   DISULFID    836    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
FT   DISULFID    871    882       BY SIMILARITY.
FT   DISULFID    876    891       BY SIMILARITY.
FT   DISULFID    893    902       BY SIMILARITY.
FT   DISULFID    909    920       BY SIMILARITY.
FT   DISULFID    914    929       BY SIMILARITY.
FT   DISULFID    931    940       BY SIMILARITY.
FT   DISULFID    947    958       BY SIMILARITY.
FT   DISULFID    952    967       BY SIMILARITY.
FT   DISULFID    969    978       BY SIMILARITY.
FT   DISULFID   1023   1034       BY SIMILARITY.
FT   DISULFID   1028   1043       BY SIMILARITY.
FT   DISULFID   1045   1054       BY SIMILARITY.
FT   DISULFID   1061   1072       BY SIMILARITY.
FT   DISULFID   1066   1081       BY SIMILARITY.
FT   DISULFID   1083   1092       BY SIMILARITY.
FT   DISULFID   1099   1120       BY SIMILARITY.
FT   DISULFID   1114   1129       BY SIMILARITY.
FT   DISULFID   1131   1140       BY SIMILARITY.
FT   DISULFID   1147   1158       BY SIMILARITY.
FT   DISULFID   1152   1167       BY SIMILARITY.
FT   DISULFID   1169   1178       BY SIMILARITY.
FT   DISULFID   1185   1196       BY SIMILARITY.
FT   DISULFID   1190   1205       BY SIMILARITY.
FT   DISULFID   1207   1216       BY SIMILARITY.
FT   DISULFID   1223   1242       BY SIMILARITY.
FT   DISULFID   1236   1251       BY SIMILARITY.
FT   DISULFID   1253   1262       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          24.0%;  Score 291;  DB 1;  Length 2437;
  Best Local Similarity  44.8%;  Pred. No. 8.36e-41;
  Matches   43;  Conservative   15;  Mismatches  31;  Indels   7;  Gaps   5;

Db    474 HCICMPGYEGVFCQINSDDCASQPCLNG-KCIDKINSFHCECPKGFSGSLCQV--DV-D- 528
          :| ||||| |  |  | |||  : | ||  |:| :||: | |: |:|| ||::   :
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    529 -E-CASTPCKNGAKCTDGPNKYTCECTPGFSGIHCE 562
           | :| | |||:| |  |:  |:| |||:|  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98
```

```
RESULT    8
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   570       FIBROPELLIN C.
FT   DOMAIN      18    55       EGF-LIKE 1.
FT   DOMAIN      62   175       CUB.
FT   DOMAIN     176   212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214   250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252   288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290   326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328   364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366   402       EGF-LIKE 7.
FT   DOMAIN     404   440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442   570       AVIDIN-LIKE.
FT   DISULFID    23    34       BY SIMILARITY.
FT   DISULFID    28    43       BY SIMILARITY.
FT   DISULFID    45    54       BY SIMILARITY.
FT   DISULFID   180   191       BY SIMILARITY.
FT   DISULFID   185   200       BY SIMILARITY.
FT   DISULFID   202   211       BY SIMILARITY.
FT   DISULFID   218   229       BY SIMILARITY.
FT   DISULFID   223   238       BY SIMILARITY.
FT   DISULFID   240   249       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   261   276       BY SIMILARITY.
FT   DISULFID   278   287       BY SIMILARITY.
FT   DISULFID   294   305       BY SIMILARITY.
FT   DISULFID   299   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   CARBOHYD    30    30       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match          23.7%;  Score 287;  DB 1;  Length 570;
  Best Local Similarity  42.1%;  Pred. No. 6.39e-40;
  Matches   40;  Conservative   18;  Mismatches  31;  Indels   6;  Gaps   2;

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYVCQCLPNYTGTHCEIS--LDA-- 369
          |:| :|::| ||  | ::|    | ||: |:| |::| | |   |:|  |||:  | |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    370 --CASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402
            | :  ||||: | : |   ||:||||: | :||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    9
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643      EXTRACELLULAR.
FT   TRANSMEM  1644   1664      POTENTIAL.
FT   DOMAIN    1665   2318      CYTOPLASMIC.
FT   DOMAIN      39   1374      34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503      3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998      6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261      PEST.
FT   DOMAIN      39     78      EGF-LIKE 1.
FT   DOMAIN      79    119      EGF-LIKE 2.
FT   DOMAIN     120    157      EGF-LIKE 3.
FT   DOMAIN     159    196      EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235      EGF-LIKE 5.
FT   DOMAIN     237    273      EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313      EGF-LIKE 7.
FT   DOMAIN     315    351      EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390      EGF-LIKE 9.
FT   DOMAIN     392    430      EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468      EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506      EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544      EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581      EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619      EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656      EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694      EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731      EGF-LIKE 18.
FT   DOMAIN     735    771      EGF-LIKE 19.
FT   DOMAIN     772    809      EGF-LIKE 20.
FT   DOMAIN     811    848      EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886      EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923      EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961      EGF-LIKE 24.
FT   DOMAIN     963    999      EGF-LIKE 25.
FT   DOMAIN    1001   1035      EGF-LIKE 26.
FT   DOMAIN    1037   1083      EGF-LIKE 27.
FT   DOMAIN    1085   1121      EGF-LIKE 28.
FT   DOMAIN    1123   1159      EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204      EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245      EGF-LIKE 31.
FT   DOMAIN    1247   1288      EGF-LIKE 32.
FT   DOMAIN    1290   1326      EGF-LIKE 33.
FT   DOMAIN    1336   1374      EGF-LIKE 34.
FT   REPEAT    1388   1428      LIN/NOTCH 1.
FT   REPEAT    1429   1467      LIN/NOTCH 2.
FT   REPEAT    1468   1503      LIN/NOTCH 3.
FT   REPEAT    1784   1816      CDC10/SWI6 1.
FT   REPEAT    1817   1865      CDC10/SWI6 2.
FT   REPEAT    1866   1898      CDC10/SWI6 3.
FT   REPEAT    1899   1932      CDC10/SWI6 4.
FT   REPEAT    1933   1965      CDC10/SWI6 5.
FT   REPEAT    1966   1998      CDC10/SWI6 6.
FT   DISULFID    43     55      BY SIMILARITY.
FT   DISULFID    49     66      BY SIMILARITY.
FT   DISULFID    68     77      BY SIMILARITY.
FT   DISULFID    83     94      BY SIMILARITY.
FT   DISULFID    88    107      BY SIMILARITY.
FT   DISULFID   109    118      BY SIMILARITY.
FT   DISULFID   124    135      BY SIMILARITY.
FT   DISULFID   129    145      BY SIMILARITY.
FT   DISULFID   147    156      BY SIMILARITY.
FT   DISULFID   163    175      BY SIMILARITY.
FT   DISULFID   169    184      BY SIMILARITY.
FT   DISULFID   186    195      BY SIMILARITY.
FT   DISULFID   202    213      BY SIMILARITY.
FT   DISULFID   207    223      BY SIMILARITY.
FT   DISULFID   225    234      BY SIMILARITY.
FT   DISULFID   241    252      BY SIMILARITY.
FT   DISULFID   246    261      BY SIMILARITY.
FT   DISULFID   263    272      BY SIMILARITY.
FT   DISULFID   279    292      BY SIMILARITY.
FT   DISULFID   286    301      BY SIMILARITY.
FT   DISULFID   303    312      BY SIMILARITY.
FT   DISULFID   319    330      BY SIMILARITY.
FT   DISULFID   324    339      BY SIMILARITY.
FT   DISULFID   341    350      BY SIMILARITY.
FT   DISULFID   356    367      BY SIMILARITY.
FT   DISULFID   361    378      BY SIMILARITY.
FT   DISULFID   380    389      BY SIMILARITY.
FT   DISULFID   396    409      BY SIMILARITY.
FT   DISULFID   403    418      BY SIMILARITY.
FT   DISULFID   420    429      BY SIMILARITY.
FT   DISULFID   436    447      BY SIMILARITY.
FT   DISULFID   441    456      BY SIMILARITY.
FT   DISULFID   458    467      BY SIMILARITY.
FT   DISULFID   474    485      BY SIMILARITY.
FT   DISULFID   479    494      BY SIMILARITY.
FT   DISULFID   496    505      BY SIMILARITY.
FT   DISULFID   512    523      BY SIMILARITY.
FT   DISULFID   517    532      BY SIMILARITY.
FT   DISULFID   534    543      BY SIMILARITY.
FT   DISULFID   550    560      BY SIMILARITY.
FT   DISULFID   555    569      BY SIMILARITY.
FT   DISULFID   571    580      BY SIMILARITY.
FT   DISULFID   587    598      BY SIMILARITY.
FT   DISULFID   592    607      BY SIMILARITY.
FT   DISULFID   609    618      BY SIMILARITY.
FT   DISULFID   625    635      BY SIMILARITY.
FT   DISULFID   630    644      BY SIMILARITY.
FT   DISULFID   646    655      BY SIMILARITY.
FT   DISULFID   662    673      BY SIMILARITY.
FT   DISULFID   667    682      BY SIMILARITY.
FT   DISULFID   684    693      BY SIMILARITY.
FT   DISULFID   700    710      BY SIMILARITY.
FT   DISULFID   705    719      BY SIMILARITY.
FT   DISULFID   721    730      BY SIMILARITY.
FT   DISULFID   739    750      BY SIMILARITY.
FT   DISULFID   744    759      BY SIMILARITY.
FT   DISULFID   761    770      BY SIMILARITY.
FT   DISULFID   776    787      BY SIMILARITY.
FT   DISULFID   781    797      BY SIMILARITY.
FT   DISULFID   799    808      BY SIMILARITY.
FT   DISULFID   815    827      BY SIMILARITY.
FT   DISULFID   821    836      BY SIMILARITY.
FT   DISULFID   838    847      BY SIMILARITY.
FT   DISULFID   854    865      BY SIMILARITY.
FT   DISULFID   859    874      BY SIMILARITY.
FT   DISULFID   876    885      BY SIMILARITY.
FT   DISULFID   892    902      BY SIMILARITY.
FT   DISULFID   897    911      BY SIMILARITY.
FT   DISULFID   913    922      BY SIMILARITY.
FT   DISULFID   929    940      BY SIMILARITY.
FT   DISULFID   934    949      BY SIMILARITY.
FT   DISULFID   951    960      BY SIMILARITY.
FT   DISULFID   967    978      BY SIMILARITY.
FT   DISULFID   972    987      BY SIMILARITY.
FT   DISULFID   989    998      BY SIMILARITY.
FT   DISULFID  1005   1016      BY SIMILARITY.
FT   DISULFID  1010   1023      BY SIMILARITY.
FT   DISULFID  1025   1034      BY SIMILARITY.
FT   DISULFID  1041   1062      BY SIMILARITY.
FT   DISULFID  1056   1071      BY SIMILARITY.
FT   DISULFID  1073   1082      BY SIMILARITY.
FT   DISULFID  1089   1100      BY SIMILARITY.
FT   DISULFID  1094   1109      BY SIMILARITY.
```

```
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           23.0%;  Score 279;  DB 1;  Length 2318;
  Best Local Similarity  41.7%;  Pred. No. 3.68e-38;
  Matches   40;  Conservative   20;  Mismatches  30;  Indels   6;  Gaps   2;

Db    456 CICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMC----QLDV-- 509
          | ||:|::|  |  : |:|:   | ||: |:| ||::|| |: |:|| :|    :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    510 DECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCER 545
            | :| |:|||:|||| :   |:|  || |  ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEK 99


RESULT  10
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2075       ANK MOTIF 6.
FT   CARBOHYD   888    888       POTENTIAL.
FT   CARBOHYD   959    959       POTENTIAL.
FT   CARBOHYD  1179   1179       POTENTIAL.
FT   CARBOHYD  1241   1241       POTENTIAL.
FT   CARBOHYD  1489   1489       POTENTIAL.
FT   CARBOHYD  1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match           22.5%;  Score 273;  DB 1;  Length 2531;
  Best Local Similarity  43.3%;  Pred. No. 7.61e-37;
  Matches   42;  Conservative   12;  Mismatches  41;  Indels   2;  Gaps   2;

Db   1169 CKCVAGYHGSNCSEEINECLSQPCQNGGTCIDLTNSYKCSCPRGTQGVHCEINVDDCHPP 1228
          | |::|| | ||||: ::|  : ||||: |:|  ||| | |: |  |  |||      :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHL-PAP 62

Db   1229 LDPASRSPKCFNNGTCVDQVGGYTCTCPPGFVGERCE 1265
            |   :  | | : ||||     | | ||| |  ||
Qy     63 KSP-CEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT  11
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
```

```
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   >2444      NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737   1757       POTENTIAL.
FT   DOMAIN    1758   >2444      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20.
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    868       EGF-LIKE 22.
FT   DOMAIN     870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     908    944       EGF-LIKE 24.
FT   DOMAIN     946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 26.
FT   DOMAIN    1022   1058       EGF-LIKE 27.
FT   DOMAIN    1060   1096       EGF-LIKE 28.
FT   DOMAIN    1098   1144       EGF-LIKE 29.
FT   DOMAIN    1146   1182       EGF-LIKE 30.
FT   DOMAIN    1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1268   1306       EGF-LIKE 33.
FT   DOMAIN    1308   1347       EGF-LIKE 34.
FT   DOMAIN    1349   1385       EGF-LIKE 35.
FT   DOMAIN    1388   1427       EGF-LIKE 36.
FT   DOMAIN    1446   1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446   1481       LIN/NOTCH 1.
FT   REPEAT    1482   1523       LIN/NOTCH 2.
FT   REPEAT    1524   1563       LIN/NOTCH 3.
FT   DOMAIN    1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT    1876   1921       ANK MOTIF 1.
FT   REPEAT    1923   1954       ANK MOTIF 2.
FT   REPEAT    1956   1987       ANK MOTIF 3.
FT   REPEAT    1990   2021       ANK MOTIF 4.
FT   REPEAT    2023   2054       ANK MOTIF 5.
FT   REPEAT    2056   2087       ANK MOTIF 6.
FT   DOMAIN    1576   1579       POLY-VAL.
FT   DOMAIN    1662   1665       POLY-ARG.
FT   DOMAIN    1729   1732       POLY-PRO.
FT   DOMAIN    1741   1744       POLY-ALA.
FT   DOMAIN    1902   1905       POLY-GLU.
FT   DOMAIN    2260   2263       POLY-GLY.
FT   DOMAIN    2404   2407       POLY-GLN.
FT   DOMAIN    2411   2418       POLY-PRO.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
FT   DISULFID   339    350       BY SIMILARITY.
FT   DISULFID   344    359       BY SIMILARITY.
FT   DISULFID   361    370       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    398       BY SIMILARITY.
FT   DISULFID   400    409       BY SIMILARITY.
FT   DISULFID   416    429       BY SIMILARITY.
FT   DISULFID   423    438       BY SIMILARITY.
FT   DISULFID   440    449       BY SIMILARITY.
FT   DISULFID   456    467       BY SIMILARITY.
FT   DISULFID   461    476       BY SIMILARITY.
FT   DISULFID   478    487       BY SIMILARITY.
FT   DISULFID   494    505       BY SIMILARITY.
FT   DISULFID   499    514       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   532    543       BY SIMILARITY.
FT   DISULFID   537    552       BY SIMILARITY.
FT   DISULFID   554    563       BY SIMILARITY.
FT   DISULFID   570    580       BY SIMILARITY.
FT   DISULFID   575    589       BY SIMILARITY.
FT   DISULFID   591    600       BY SIMILARITY.
FT   DISULFID   607    618       BY SIMILARITY.
FT   DISULFID   612    627       BY SIMILARITY.
FT   DISULFID   629    638       BY SIMILARITY.
FT   DISULFID   645    655       BY SIMILARITY.
```

```
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    867       BY SIMILARITY.
FT   DISULFID    874    885       BY SIMILARITY.
FT   DISULFID    879    894       BY SIMILARITY.
FT   DISULFID    896    905       BY SIMILARITY.
FT   DISULFID    912    923       BY SIMILARITY.
FT   DISULFID    917    932       BY SIMILARITY.
FT   DISULFID    934    943       BY SIMILARITY.
FT   DISULFID    988    999       BY SIMILARITY.
FT   DISULFID    993   1008       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1026   1037       BY SIMILARITY.
FT   DISULFID   1031   1046       BY SIMILARITY.
FT   DISULFID   1048   1057       BY SIMILARITY.
FT   DISULFID   1064   1075       BY SIMILARITY.
FT   DISULFID   1069   1084       BY SIMILARITY.
FT   DISULFID   1086   1095       BY SIMILARITY.
FT   DISULFID   1102   1123       BY SIMILARITY.
FT   DISULFID   1117   1132       BY SIMILARITY.
FT   DISULFID   1134   1143       BY SIMILARITY.
FT   DISULFID   1150   1161       BY SIMILARITY.
FT   DISULFID   1155   1170       BY SIMILARITY.
FT   DISULFID   1172   1181       BY SIMILARITY.
FT   DISULFID   1188   1199       BY SIMILARITY.
FT   DISULFID   1193   1208       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          22.4%;  Score 272;  DB 1;  Length 2444;
  Best Local Similarity  44.3%;  Pred. No. 1.26e-36;
  Matches   43;  Conservative   11;  Mismatches  41;  Indels   2;  Gaps   2;

Db   1170 CKCVAGYHGVNCSEEIDECLSHPCQNGGTCLDLPNTYKCSCPRGTQGVHCEINVDDCNPP 1229
          | |::|| | ||||: |:|  | ||||: |:|  |:| | |: |  |  |||     :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHL-PAP 62

Db   1230 VDPVSRSPKCFNNGTCVDQVGGYSCTCPPGFVGERCE 1266
           |   :  | | : ||||     | | ||| |  ||
Qy     63 KSPCE-GTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   12
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
FT   DOMAIN     869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907    943       EGF-LIKE 24.
FT   DOMAIN     945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983   1019       EGF-LIKE 26.
FT   DOMAIN    1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059   1095       EGF-LIKE 28.
FT   DOMAIN    1097   1143       EGF-LIKE 29.
FT   DOMAIN    1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267   1305       EGF-LIKE 33.
FT   DOMAIN    1307   1346       EGF-LIKE 34.
FT   DOMAIN    1348   1384       EGF-LIKE 35.
FT   DOMAIN    1387   1426       EGF-LIKE 36.
```

```
FT   DOMAIN     1449   1462       CYS-RICH.
FT   DOMAIN     1865   2076       6 X ANK MOTIF REPEATS.
FT   REPEAT     1865   1910       ANK MOTIF 1.
FT   REPEAT     1912   1942       ANK MOTIF 2.
FT   REPEAT     1944   1975       ANK MOTIF 3.
FT   REPEAT     1978   2009       ANK MOTIF 4.
FT   REPEAT     2011   2042       ANK MOTIF 5.
FT   REPEAT     2044   2076       ANK MOTIF 6.
FT   DISULFID     24     37       BY SIMILARITY.
FT   DISULFID     31     46       BY SIMILARITY.
FT   DISULFID     48     57       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    261    272       BY SIMILARITY.
FT   DISULFID    266    281       BY SIMILARITY.
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    299    312       BY SIMILARITY.
FT   DISULFID    306    321       BY SIMILARITY.
FT   DISULFID    323    332       BY SIMILARITY.
FT   DISULFID    339    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    361    370       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    398       BY SIMILARITY.
FT   DISULFID    400    409       BY SIMILARITY.
FT   DISULFID    416    429       BY SIMILARITY.
FT   DISULFID    423    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    456    467       BY SIMILARITY.
FT   DISULFID    461    476       BY SIMILARITY.
FT   DISULFID    478    487       BY SIMILARITY.
FT   DISULFID    494    505       BY SIMILARITY.
FT   DISULFID    499    514       BY SIMILARITY.
FT   DISULFID    516    525       BY SIMILARITY.
FT   DISULFID    532    543       BY SIMILARITY.
FT   DISULFID    537    552       BY SIMILARITY.
FT   DISULFID    554    563       BY SIMILARITY.
FT   DISULFID    570    580       BY SIMILARITY.
FT   DISULFID    575    589       BY SIMILARITY.
FT   DISULFID    591    600       BY SIMILARITY.
FT   DISULFID    607    618       BY SIMILARITY.
FT   DISULFID    612    627       BY SIMILARITY.
FT   DISULFID    629    638       BY SIMILARITY.
FT   DISULFID    645    655       BY SIMILARITY.
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    866       BY SIMILARITY.
FT   DISULFID    873    884       BY SIMILARITY.
FT   DISULFID    878    893       BY SIMILARITY.
FT   DISULFID    895    904       BY SIMILARITY.
FT   DISULFID    911    922       BY SIMILARITY.
FT   DISULFID    916    931       BY SIMILARITY.
FT   DISULFID    933    942       BY SIMILARITY.
FT   DISULFID    987    998       BY SIMILARITY.
FT   DISULFID    992   1007       BY SIMILARITY.
FT   DISULFID   1009   1018       BY SIMILARITY.
FT   DISULFID   1025   1036       BY SIMILARITY.
FT   DISULFID   1030   1045       BY SIMILARITY.
FT   DISULFID   1047   1056       BY SIMILARITY.
FT   DISULFID   1063   1074       BY SIMILARITY.
FT   DISULFID   1068   1083       BY SIMILARITY.
FT   DISULFID   1085   1094       BY SIMILARITY.
FT   DISULFID   1101   1122       BY SIMILARITY.
FT   DISULFID   1116   1131       BY SIMILARITY.
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.
FT   DISULFID   1225   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1271   1284       BY SIMILARITY.
FT   DISULFID   1276   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1311   1322       BY SIMILARITY.
FT   DISULFID   1316   1334       BY SIMILARITY.
FT   DISULFID   1336   1345       BY SIMILARITY.
FT   DISULFID   1352   1363       BY SIMILARITY.
FT   DISULFID   1357   1372       BY SIMILARITY.
FT   DISULFID   1374   1383       BY SIMILARITY.
FT   DISULFID   1391   1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          22.4%;  Score 272;  DB 1;  Length 2531;
  Best Local Similarity  42.1%;  Pred. No. 1.26e-36;
  Matches   40;  Conservative   18;  Mismatches  31;  Indels   6;  Gaps   4;

Db    931 CDCLPGFQGAFCEEDINECATNPCQNGANCTDCVDSYTCTCPTGFNGIHCE---N-NTP- 985
          |:|:||: |  | |: ::|  : ||||| | | |:||:| |: |::|  ||   :  :|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    986 D-CTESSCFNGGTCVDGINSFTCLCPPGFTGSYCQ 1019
            |  : | ||: |||  |   | | ||| |: |:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT  13
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. VIROL. 66:2594-2599(1992).
```

```
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG, AND TESTIS;
RX   MEDLINE; 96281668.
RA   UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   DEVELOPMENT 122:2251-2259(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; G1714084; -.
DR   EMBL; U43691; G1401160; -.
DR   PIR; A38072; TVMVT3.
DR   MGD; MGI:107471; NOTCH4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   DOMAIN      21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1444   1464       POTENTIAL.
FT   DOMAIN    1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     60       EGF-LIKE 1.
FT   DOMAIN      61    112       EGF-LIKE 2.
FT   DOMAIN     115    152       EGF-LIKE 3.
FT   DOMAIN     153    189       EGF-LIKE 4.
FT   DOMAIN     191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     231    271       EGF-LIKE 6.
FT   DOMAIN     273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     389    427       EGF-LIKE 10.
FT   DOMAIN     429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     622    656       EGF-LIKE 16.
FT   DOMAIN     658    686       EGF-LIKE 17.
FT   DOMAIN     688    724       EGF-LIKE 18.
FT   DOMAIN     726    762       EGF-LIKE 19.
FT   DOMAIN     764    800       EGF-LIKE 20.
FT   DOMAIN     803    839       EGF-LIKE 21.
FT   DOMAIN     841    877       EGF-LIKE 22.
FT   DOMAIN     878    924       EGF-LIKE 23.
FT   DOMAIN     926    962       EGF-LIKE 24.
FT   DOMAIN     964   1000       EGF-LIKE 25.
FT   DOMAIN    1002   1040       EGF-LIKE 26.
FT   DOMAIN    1042   1081       EGF-LIKE 27.
FT   DOMAIN    1083   1122       EGF-LIKE 28.
FT   DOMAIN    1126   1167       EGF-LIKE 29.
FT   DOMAIN    1168   1282       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1168   1208       LIN/NOTCH 1.
FT   REPEAT    1209   1242       LIN/NOTCH 2.
FT   REPEAT    1243   1282       LIN/NOTCH 3.
FT   DOMAIN    1572   1785       6 X ANK MOTIF REPEATS.
FT   REPEAT    1572   1603       ANK MOTIF 1.
FT   REPEAT    1622   1653       ANK MOTIF 2.
FT   REPEAT    1654   1685       ANK MOTIF 3.
FT   REPEAT    1688   1719       ANK MOTIF 4.
FT   REPEAT    1721   1752       ANK MOTIF 5.
FT   REPEAT    1754   1785       ANK MOTIF 6.
FT   DISULFID    25     38       BY SIMILARITY.
FT   DISULFID    32     48       BY SIMILARITY.
FT   DISULFID    50     59       BY SIMILARITY.
FT   DISULFID    65     77       BY SIMILARITY.
FT   DISULFID    71    100       BY SIMILARITY.
FT   DISULFID   102    111       BY SIMILARITY.
FT   DISULFID   119    130       BY SIMILARITY.
FT   DISULFID   124    140       BY SIMILARITY.
FT   DISULFID   142    151       BY SIMILARITY.
FT   DISULFID   157    168       BY SIMILARITY.
FT   DISULFID   162    177       BY SIMILARITY.
FT   DISULFID   179    188       BY SIMILARITY.
FT   DISULFID   195    208       BY SIMILARITY.
FT   DISULFID   202    217       BY SIMILARITY.
FT   DISULFID   219    228       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   277    288       BY SIMILARITY.
FT   DISULFID   282    297       BY SIMILARITY.
FT   DISULFID   299    308       BY SIMILARITY.
FT   DISULFID   315    329       BY SIMILARITY.
FT   DISULFID   323    338       BY SIMILARITY.
FT   DISULFID   340    349       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    376       BY SIMILARITY.
FT   DISULFID   378    387       BY SIMILARITY.
FT   DISULFID   393    404       BY SIMILARITY.
FT   DISULFID   398    415       BY SIMILARITY.
FT   DISULFID   417    426       BY SIMILARITY.
FT   DISULFID   433    449       BY SIMILARITY.
FT   DISULFID   443    458       BY SIMILARITY.
FT   DISULFID   460    469       BY SIMILARITY.
FT   DISULFID   476    487       BY SIMILARITY.
FT   DISULFID   481    496       BY SIMILARITY.
FT   DISULFID   498    507       BY SIMILARITY.
FT   DISULFID   514    525       BY SIMILARITY.
FT   DISULFID   519    534       BY SIMILARITY.
FT   DISULFID   536    545       BY SIMILARITY.
FT   DISULFID   552    563       BY SIMILARITY.
FT   DISULFID   557    572       BY SIMILARITY.
FT   DISULFID   574    583       BY SIMILARITY.
FT   DISULFID   590    601       BY SIMILARITY.
FT   DISULFID   595    610       BY SIMILARITY.
FT   DISULFID   612    621       BY SIMILARITY.
FT   DISULFID   626    637       BY SIMILARITY.
FT   DISULFID   631    646       BY SIMILARITY.
FT   DISULFID   648    655       BY SIMILARITY.
FT   DISULFID   662    669       BY SIMILARITY.
FT   DISULFID   664    674       BY SIMILARITY.
FT   DISULFID   676    685       BY SIMILARITY.
FT   DISULFID   692    703       BY SIMILARITY.
FT   DISULFID   697    712       BY SIMILARITY.
FT   DISULFID   714    723       BY SIMILARITY.
FT   DISULFID   730    741       BY SIMILARITY.
FT   DISULFID   735    750       BY SIMILARITY.
FT   DISULFID   752    761       BY SIMILARITY.
```

```
FT   DISULFID    768    779       BY SIMILARITY.
FT   DISULFID    773    788       BY SIMILARITY.
FT   DISULFID    790    799       BY SIMILARITY.
FT   DISULFID    807    818       BY SIMILARITY.
FT   DISULFID    812    827       BY SIMILARITY.
FT   DISULFID    829    838       BY SIMILARITY.
FT   DISULFID    845    856       BY SIMILARITY.
FT   DISULFID    850    865       BY SIMILARITY.
FT   DISULFID    867    876       BY SIMILARITY.
FT   DISULFID    882    903       BY SIMILARITY.
FT   DISULFID    897    912       BY SIMILARITY.
FT   DISULFID    914    923       BY SIMILARITY.
FT   DISULFID    930    941       BY SIMILARITY.
FT   DISULFID    935    950       BY SIMILARITY.
FT   DISULFID    952    961       BY SIMILARITY.
FT   DISULFID    968    979       BY SIMILARITY.
FT   DISULFID    973    988       BY SIMILARITY.
FT   DISULFID    990    999       BY SIMILARITY.
FT   DISULFID   1006   1019       BY SIMILARITY.
FT   DISULFID   1011   1028       BY SIMILARITY.
FT   DISULFID   1030   1039       BY SIMILARITY.
FT   DISULFID   1046   1057       BY SIMILARITY.
FT   DISULFID   1051   1069       BY SIMILARITY.
FT   DISULFID   1071   1080       BY SIMILARITY.
FT   DISULFID   1087   1098       BY SIMILARITY.
FT   DISULFID   1092   1110       BY SIMILARITY.
FT   DISULFID   1112   1121       BY SIMILARITY.
FT   DISULFID   1130   1142       BY SIMILARITY.
FT   DISULFID   1136   1155       BY SIMILARITY.
FT   DISULFID   1157   1166       BY SIMILARITY.
FT   CARBOHYD    711    711       POTENTIAL.
FT   CARBOHYD    960    960       POTENTIAL.
FT   CARBOHYD   1139   1139       POTENTIAL.
FT   CONFLICT     43     43       Q -> R (IN REF. 3).
FT   CONFLICT    298    298       L -> P (IN REF. 3).
FT   CONFLICT    884    884       M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match           22.2%;  Score 269;  DB 1;  Length 1964;
  Best Local Similarity  38.0%;  Pred. No. 5.69e-36;
  Matches   38;  Conservative   22;  Mismatches  37;  Indels   3;  Gaps   3;

Db    825 ARCLCSPGYTGSSCQTLIDLCARKPCPHTARCLQSGPSFQCLCLQGWTGALCDFPLSCQM 884
          :|| | |||:| :|    | |    | : |:|::   |: ||| :| :| ||::| : ::
Qy      2 PRCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIP-P-HL 59

Db    885 AAMSQGIEISGLCQNGGLCIDTGSSYFCRCPPGFQGKLCQ 924
          :|     | :  ||||: |:| |:   |:| ||| |  |:
Qy     60 PAPKSPCEGTE-CQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   14
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     44       POTENTIAL.
FT   CHAIN        45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN       45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1746   1766       POTENTIAL.
```

```
FT   DOMAIN     1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN       58     95       EGF-LIKE 1.
FT   DOMAIN       96    136       EGF-LIKE 2.
FT   DOMAIN      139    176       EGF-LIKE 3.
FT   DOMAIN      177    215       EGF-LIKE 4.
FT   DOMAIN      217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      255    291       EGF-LIKE 6.
FT   DOMAIN      293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      409    447       EGF-LIKE 10.
FT   DOMAIN      449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      867    905       EGF-LIKE 22.
FT   DOMAIN      907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984   1020       EGF-LIKE 25.
FT   DOMAIN     1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1060   1096       EGF-LIKE 27.
FT   DOMAIN     1098   1134       EGF-LIKE 28.
FT   DOMAIN     1136   1181       EGF-LIKE 29.
FT   DOMAIN     1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1297   1335       EGF-LIKE 33.
FT   DOMAIN     1337   1373       EGF-LIKE 34.
FT   DOMAIN     1375   1412       EGF-LIKE 35.
FT   DOMAIN     1415   1451       EGF-LIKE 36.
FT   DOMAIN     1475   1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1475   1513       LIN/NOTCH 1.
FT   REPEAT     1514   1553       LIN/NOTCH 2.
FT   REPEAT     1554   1593       LIN/NOTCH 3.
FT   DOMAIN     1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN     2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID     62     73       BY SIMILARITY.
FT   DISULFID     67     83       BY SIMILARITY.
FT   DISULFID     85     94       BY SIMILARITY.
FT   DISULFID    100    111       BY SIMILARITY.
FT   DISULFID    105    124       BY SIMILARITY.
FT   DISULFID    126    135       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    181    192       BY SIMILARITY.
FT   DISULFID    186    203       BY SIMILARITY.
FT   DISULFID    205    214       BY SIMILARITY.
FT   DISULFID    221    232       BY SIMILARITY.
FT   DISULFID    226    241       BY SIMILARITY.
FT   DISULFID    243    252       BY SIMILARITY.
FT   DISULFID    259    270       BY SIMILARITY.
FT   DISULFID    264    279       BY SIMILARITY.
FT   DISULFID    281    290       BY SIMILARITY.
FT   DISULFID    297    308       BY SIMILARITY.
FT   DISULFID    302    317       BY SIMILARITY.
FT   DISULFID    319    328       BY SIMILARITY.
FT   DISULFID    335    349       BY SIMILARITY.
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    396       BY SIMILARITY.
FT   DISULFID    398    407       BY SIMILARITY.
FT   DISULFID    413    424       BY SIMILARITY.
FT   DISULFID    418    435       BY SIMILARITY.
FT   DISULFID    437    446       BY SIMILARITY.
FT   DISULFID    453    465       BY SIMILARITY.
FT   DISULFID    459    474       BY SIMILARITY.
FT   DISULFID    476    485       BY SIMILARITY.
FT   DISULFID    492    503       BY SIMILARITY.
FT   DISULFID    497    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    579       BY SIMILARITY.
FT   DISULFID    573    588       BY SIMILARITY.
FT   DISULFID    590    599       BY SIMILARITY.
FT   DISULFID    606    616       BY SIMILARITY.
FT   DISULFID    611    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    654       BY SIMILARITY.
FT   DISULFID    648    663       BY SIMILARITY.
FT   DISULFID    665    674       BY SIMILARITY.
FT   DISULFID    681    692       BY SIMILARITY.
FT   DISULFID    686    701       BY SIMILARITY.
FT   DISULFID    703    712       BY SIMILARITY.
FT   DISULFID    719    730       BY SIMILARITY.
FT   DISULFID    724    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          22.2%;  Score 269;  DB 1;  Length 2703;
  Best Local Similarity  41.5%;  Pred. No. 5.69e-36;
  Matches   39;  Conservative   17;  Mismatches  32;  Indels   6;  Gaps   4;

Db   1008 CTCPLGFSGINCQTNDEDCTESSCLNGGSCIDGINGYNCSCLAGYSGANCQY--KL-N-K 1063
            | |  |::| ||  |::|| :  | ||: |:| :|:|:| |  ||||  |:    |  |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db   1064 --CDSNPCLNGATCHEQNNEYTCHCPSGFTGKQC 1095
             |::  | ||| | :| |  |:| :|| | :|
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPEC 97


RESULT   15
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
```

```
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
     HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    19       POTENTIAL.
FT   CHAIN       20  2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20  1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729  1750       POTENTIAL.
FT   DOMAIN    1751  2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    57       EGF-LIKE 1.
FT   DOMAIN      58    99       EGF-LIKE 2.
FT   DOMAIN     102   140       EGF-LIKE 3.
FT   DOMAIN     141   177       EGF-LIKE 4.
FT   DOMAIN     179   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217   254       EGF-LIKE 6.
FT   DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371   409       EGF-LIKE 10.
FT   DOMAIN     411   449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451   487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489   525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527   563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565   600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602   638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     640   675       EGF-LIKE 17.
FT   DOMAIN     677   713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715   750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     752   788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     790   826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     828   866       EGF-LIKE 22.
FT   DOMAIN     868   904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     906   942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     944   980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     982  1018       EGF-LIKE 26.
FT   DOMAIN    1020  1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1058  1094       EGF-LIKE 28.
FT   DOMAIN    1096  1142       EGF-LIKE 29.
FT   DOMAIN    1144  1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1182  1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1220  1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1266  1304       EGF-LIKE 33.
FT   DOMAIN    1306  1346       EGF-LIKE 34.
FT   DOMAIN    1347  1383       EGF-LIKE 35.
FT   DOMAIN    1386  1424       EGF-LIKE 36.
FT   DOMAIN    1441  1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1441  1478       LIN/NOTCH 1.
FT   REPEAT    1479  1520       LIN/NOTCH 2.
FT   REPEAT    1521  1560       LIN/NOTCH 3.
FT   DOMAIN    1871  2083       6 X ANK MOTIF REPEATS.
FT   DISULFID    22    35       BY SIMILARITY.
FT   DISULFID    29    45       BY SIMILARITY.
FT   DISULFID    47    56       BY SIMILARITY.
FT   DISULFID    62    74       BY SIMILARITY.
FT   DISULFID    68    87       BY SIMILARITY.
FT   DISULFID    89    98       BY SIMILARITY.
FT   DISULFID   106   117       BY SIMILARITY.
FT   DISULFID   111   128       BY SIMILARITY.
FT   DISULFID   130   139       BY SIMILARITY.
FT   DISULFID   145   156       BY SIMILARITY.
FT   DISULFID   150   165       BY SIMILARITY.
FT   DISULFID   167   176       BY SIMILARITY.
FT   DISULFID   183   194       BY SIMILARITY.
FT   DISULFID   188   203       BY SIMILARITY.
FT   DISULFID   205   214       BY SIMILARITY.
FT   DISULFID   221   232       BY SIMILARITY.
FT   DISULFID   226   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   265   280       BY SIMILARITY.
FT   DISULFID   282   291       BY SIMILARITY.
FT   DISULFID   298   311       BY SIMILARITY.
FT   DISULFID   305   320       BY SIMILARITY.
FT   DISULFID   322   331       BY SIMILARITY.
FT   DISULFID   338   349       BY SIMILARITY.
FT   DISULFID   343   358       BY SIMILARITY.
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   375   386       BY SIMILARITY.
FT   DISULFID   380   397       BY SIMILARITY.
FT   DISULFID   399   408       BY SIMILARITY.
FT   DISULFID   415   428       BY SIMILARITY.
FT   DISULFID   422   437       BY SIMILARITY.
FT   DISULFID   439   448       BY SIMILARITY.
FT   DISULFID   455   466       BY SIMILARITY.
FT   DISULFID   460   475       BY SIMILARITY.
FT   DISULFID   477   486       BY SIMILARITY.
FT   DISULFID   493   504       BY SIMILARITY.
FT   DISULFID   498   513       BY SIMILARITY.
FT   DISULFID   515   524       BY SIMILARITY.
FT   DISULFID   531   542       BY SIMILARITY.
FT   DISULFID   536   551       BY SIMILARITY.
FT   DISULFID   553   562       BY SIMILARITY.
FT   DISULFID   569   579       BY SIMILARITY.
FT   DISULFID   574   588       BY SIMILARITY.
FT   DISULFID   590   599       BY SIMILARITY.
FT   DISULFID   606   617       BY SIMILARITY.
FT   DISULFID   611   626       BY SIMILARITY.
FT   DISULFID   628   637       BY SIMILARITY.
FT   DISULFID   644   654       BY SIMILARITY.
FT   DISULFID   649   663       BY SIMILARITY.
FT   DISULFID   665   674       BY SIMILARITY.
FT   DISULFID   681   692       BY SIMILARITY.
FT   DISULFID   686   701       BY SIMILARITY.
FT   DISULFID   703   712       BY SIMILARITY.
FT   DISULFID   719   729       BY SIMILARITY.
FT   DISULFID   724   738       BY SIMILARITY.
FT   DISULFID   740   749       BY SIMILARITY.
FT   DISULFID   756   767       BY SIMILARITY.
FT   DISULFID   761   776       BY SIMILARITY.
FT   DISULFID   778   787       BY SIMILARITY.
FT   DISULFID   794   805       BY SIMILARITY.
FT   DISULFID   799   814       BY SIMILARITY.
FT   DISULFID   816   825       BY SIMILARITY.
FT   DISULFID   832   843       BY SIMILARITY.
FT   DISULFID   837   854       BY SIMILARITY.
FT   DISULFID   856   865       BY SIMILARITY.
FT   DISULFID   872   883       BY SIMILARITY.
FT   DISULFID   877   892       BY SIMILARITY.
FT   DISULFID   894   903       BY SIMILARITY.
FT   DISULFID   910   921       BY SIMILARITY.
FT   DISULFID   915   930       BY SIMILARITY.
FT   DISULFID   932   941       BY SIMILARITY.
FT   DISULFID   986   997       BY SIMILARITY.
FT   DISULFID   991  1006       BY SIMILARITY.
FT   DISULFID  1008  1017       BY SIMILARITY.
```

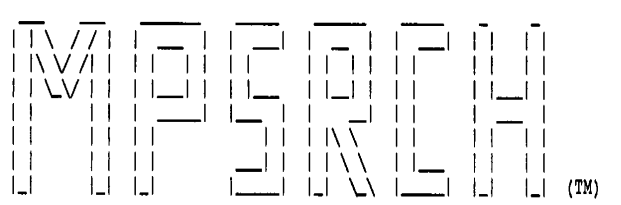
```
FT   DISULFID   1024   1035       BY SIMILARITY.
FT   DISULFID   1029   1044       BY SIMILARITY.
FT   DISULFID   1046   1055       BY SIMILARITY.
FT   DISULFID   1062   1073       BY SIMILARITY.
FT   DISULFID   1067   1082       BY SIMILARITY.
FT   DISULFID   1084   1093       BY SIMILARITY.
FT   DISULFID   1100   1121       BY SIMILARITY.
FT   DISULFID   1115   1130       BY SIMILARITY.
FT   DISULFID   1132   1141       BY SIMILARITY.
FT   DISULFID   1148   1159       BY SIMILARITY.
FT   DISULFID   1153   1168       BY SIMILARITY.
FT   DISULFID   1170   1179       BY SIMILARITY.
FT   DISULFID   1186   1197       BY SIMILARITY.
FT   DISULFID   1191   1206       BY SIMILARITY.
FT   DISULFID   1208   1217       BY SIMILARITY.
FT   DISULFID   1224   1243       BY SIMILARITY.
FT   DISULFID   1237   1252       BY SIMILARITY.
FT   DISULFID   1254   1263       BY SIMILARITY.
FT   DISULFID   1270   1283       BY SIMILARITY.
FT   DISULFID   1275   1292       BY SIMILARITY.
FT   DISULFID   1294   1303       BY SIMILARITY.
FT   DISULFID   1310   1321       BY SIMILARITY.
FT   DISULFID   1315   1333       BY SIMILARITY.
FT   DISULFID   1335   1344       BY SIMILARITY.
FT   DISULFID   1351   1362       BY SIMILARITY.
FT   DISULFID   1356   1371       BY SIMILARITY.
FT   DISULFID   1373   1382       BY SIMILARITY.
FT   DISULFID   1390   1401       BY SIMILARITY.
FT   DISULFID   1395   1412       BY SIMILARITY.
FT   DISULFID   1414   1423       BY SIMILARITY.
FT   CARBOHYD    462    462       POTENTIAL.
FT   CARBOHYD    887    887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          22.1%;  Score 268;  DB 1;  Length 2524;
  Best Local Similarity  41.7%;  Pred. No. 9.41e-36;
  Matches   40;  Conservative   14;  Mismatches  36;  Indels   6;  Gaps   3;

Db    474 QCICMPGYEGLYCETNIDECASNPCLHNGKCIDKINEFRCDCPTGFSGNLCQ---H--DF 528
          :| ||||| |  |  | |:|  : | : : |:| :| : | |: |:|| ||:   |
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    529 DE-CTSTPCKNGAKCLDGPNSYTCQCTEGFTGRHCE 563
            | :| | |||:|:|  |   |||  || |  ||
          63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98

Search completed: Fri May 28 08:54:52 1999
Job time : 22 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:55:11 1999; MasPar time 12.95 Seconds
674.489 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-5
Description: (1-160) from US09191647.pep
Perfect Score: 1212
Sequence: 1 WPRCECMPGYAGDNCSENQD.....NGGNDHIAVXLYXGHVRFSY 160

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 40.449; Variance 82.697; scale 0.489

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1039	85.7	1531	11	O88279	MEGF4.	3.84e-182
2	684	56.4	1523	11	O88280	MEGF5.	3.92e-110
3	669	55.2	739	4	O75094	MEGF5 (FRAGMENT).	3.98e-107
4	491	40.5	530	5	Q24526	SLIT LOCUS ENCODING A	7.89e-72
5	480	39.6	601	5	Q20204	F40E10.4 PROTEIN (FRAG	1.12e-69
6	322	26.6	721	13	Q91902	X-DELTA-1.	2.37e-39
7	309	25.5	406	5	Q25059	FIBROPELLIN III (FRAGM	6.34e-37
8	304	25.1	728	13	Q90656	TRANSMEMBRANE PROTEIN	5.40e-36
9	303	25.0	1476	13	Q90285	PUTATIVE EXTRACELLULAR	8.28e-36
10	301	24.8	529	5	Q25058	FIBROPELLIN IA (FRAGME	1.95e-35
11	298	24.6	717	13	P87357	DELTAD TRANSMEMBRANE P	7.00e-35
12	297	24.5	3313	11	O88278	MEGF2.	1.07e-34
13	294	24.3	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	3.85e-34
14	295	24.3	2408	4	Q92566	MYELOBLAST KIAA0279 (F	2.52e-34
15	294	24.3	2470	11	O35516	CELL SURFACE PROTEIN.	3.85e-34
16	291	24.0	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.38e-33
17	285	23.5	802	13	O57462	DELTAA.	1.76e-32
18	284	23.4	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	2.69e-32
19	283	23.3	752	13	O42374	NOTCH RECEPTOR PROTEIN	4.11e-32
20	283	23.3	1218	4	O14902	TRANSMEMBRANE PROTEIN	4.11e-32
21	283	23.3	1218	4	Q15816	TRANSMEMBRANE PROTEIN	4.11e-32
22	283	23.3	1219	11	Q63722	JAGGED PROTEIN.	4.11e-32
23	283	23.3	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	4.11e-32
24	283	23.3	2352	5	O61240	HRNOTCH PROTEIN.	4.11e-32
25	281	23.2	1218	4	O15122	JAGGED1.	9.58e-32
26	278	22.9	1372	5	P91526	SIMILARITY TO MULTIPLE	3.40e-31
27	278	22.9	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	3.40e-31
28	275	22.7	2447	13	O13149	NOTCH 2 (FRAGMENT).	1.21e-30
29	273	22.5	387	11	Q06007	NOTCH PROTEIN HOMOLOG	2.80e-30
30	273	22.5	2531	5	O16004	NOTCH HOMOLOG.	2.80e-30
31	272	22.4	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	4.26e-30
32	266	21.9	615	13	O57409	DELTAB.	5.30e-29
33	265	21.9	1964	11	O35442	NOTCH4.	8.05e-29
34	262	21.6	955	4	Q99466	NOTCH4 (FRAGMENT).	2.83e-28
35	262	21.6	1999	4	Q99940	NOTCH4.	2.83e-28
36	262	21.6	2003	4	O00306	NOTCH4.	2.83e-28
37	256	21.1	642	13	P79941	NOTCH LIGAND X-DELTA-2	3.45e-27
38	255	21.0	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	5.24e-27
39	248	20.5	1202	11	P97607	JAGGED2 (FRAGMENT).	9.58e-26
40	245	20.2	434	11	O55139	JAGGED2 PROTEIN (FRAGM	3.31e-25
41	245	20.2	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	3.31e-25
42	244	20.1	592	11	O88516	DELTA-LIKE 3 ALTERNATE	5.01e-25
43	242	20.0	585	11	O35675	M-DELTA-LIKE 3 GENE PR	1.14e-24
44	237	19.6	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	8.94e-24
45	235	19.4	156	5	Q26661	EPIDERMAL GROWTH FACTO	2.03e-23

ALIGNMENTS

RESULT 1
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 85.7%; Score 1039; DB 11; Length 1531;
Best Local Similarity 86.2%; Pred. No. 3.84e-182;
Matches 137; Conservative 11; Mismatches 8; Indels 3; Gaps 1;

```
Db   1067 PRCECVPGYTGDNCSKNQDDCKDHQCQNGAQCVDEINSYACLCAEGYSGQLCEIPP---A 1123
          |||||:|||:||||| |||||:||:|||||||:||:|||:||||||||||||||||   |
Qy      2 PRCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA 61

Db   1124 PRNSCEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWP 1183
          |:::|||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Qy     62 PKSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWX 121

Db   1184 RANITLQVSTAEDNGILLYNGDNDHIAVELYQGHVRVSY 1222
          | |||||| |||||||||||| |||||| || |||| ||
Qy    122 RXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160
```

RESULT 2

```
ID   O88280      PRELIMINARY;      PRT;  1523 AA.
AC   O88280;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5.
GN   MEGF5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011531; D1033424; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1523 AA;  167767 MW;  2BD845D0 CRC32;

  Query Match           56.4%;  Score 684;  DB 11;  Length 1523;
  Best Local Similarity  52.2%;  Pred. No. 3.92e-110;
  Matches   83;  Conservative   34;  Mismatches  41;  Indels   1;  Gaps   1;

Db   1059 RCECVPGYSGKLCETDNDDCVAHKCRHGAQCVDAVNGYTCICPQGFSGLFCEHPPPMVLL 1118
           ||||:|||:|  |  : |||  |:|::||||:| ||:|:|:|::|:|| :|| || :
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA- 61

Db   1119 QTSPCDQYECQNGAQCIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRP 1178
            |||:  |||||| |:     | |:| |||:|| ||||::|||| :|:|:::    :
Qy     62 PKSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWX 121

Db   1179 QANISLQVATDKDNGILLYKGDNDPLALELYQGHVRLVY 1217
           : ||:||| |  |||||||:| || :|: || ||||: |
Qy    122 RXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT   3
ID   O75094      PRELIMINARY;      PRT;  739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;

  Query Match           55.2%;  Score 669;  DB 4;  Length 739;
  Best Local Similarity  51.3%;  Pred. No. 3.98e-107;
  Matches   81;  Conservative   34;  Mismatches  42;  Indels   1;  Gaps   1;

Db    276 CECVPGYSGKLCETDNDDCVAHKCRHGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVLLQ 335
            |||:|||:|  |  : |||  |:|::||||:| :|:|:|  |::|:|| :|| || :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-P 62

Db    336 TSPCDQYECQNGAQCIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRPQ 395
            |||:  |||||| |:     | |:| |||:|| ||||::|||| :|:|:::    : :
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWXR 122

Db    396 ANISLQVATDKDNGILLYKGDNDPLALELYQGHVRLVY 433
           ||:||| |  |||||||:| || :|: || ||||: |
Qy    123 XNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT   4
ID   Q24526      PRELIMINARY;      PRT;  530 AA.
AC   Q24526;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SLIT LOCUS ENCODING A PROTEIN ASSOCIATED WITH NEURAL DEVELOPMENT WITH
DE   52D EGF HOMOLOGOUS DOMAINS (FRAGMENT).
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 89077533.
RA   ROTHBERG J.M., HARTLEY D.A., WALTHER Z., ARTAVANIS-TSAKONAS S.;
RT   "slit: an EGF-homologous locus of D. melanogaster involved in the
RT   development of the embryonic central nervous system.";
RL   CELL 55:1047-1059(1988).
DR   EMBL; M23543; G514357; -.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
KW   NEUROGENESIS; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     530    530
SQ   SEQUENCE   530 AA;  59457 MW;  10E5764D CRC32;

  Query Match           40.5%;  Score 491;  DB 5;  Length 530;
  Best Local Similarity  39.5%;  Pred. No. 7.89e-72;
  Matches   64;  Conservative   43;  Mismatches  49;  Indels   6;  Gaps   4;

Db    170 CDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMY 229
            |:| :|: | ||::| |||::| |||||: |:| :| | | |:: |:|  ||    ::
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPA-- 61

Db    230 PQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLR 289
            | |||:  || :|      ||:  :|:| ||: |  || | |::||  ::::::  |:
Qy     62 PK-SPCEGTECQNGA--NCVDQGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db    290 TRPEANVTI-VFSSGQNGILMYDGQDAHLAVELFNGRIRVSY 330
            :  |:|: ||:: :||||:|:| : |:|| |: |::| ||
Qy    119 NWXRXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT   5
ID   Q20204      PRELIMINARY;      PRT;  601 AA.
AC   Q20204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F40E10.4 PROTEIN (FRAGMENT).
GN   F40E10.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   SMYE R.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z69792; E1346469; -.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match           39.6%;  Score 480;  DB 5;  Length 601;
  Best Local Similarity  40.1%;  Pred. No. 1.12e-69;
  Matches    65;  Conservative   33;  Mismatches  56;  Indels   8;  Gaps   8;

Db    202 CMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEY 261
          | | ||::|:||  | |||::  ||||: |:| : || |||  ||:|| |||||  :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLP-A- 61

Db    262 QKTDACQQSACGQG-ECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPLA 320
           |: :|: : |  | :||  | :    | |  ||:|| |:: :|| | :  :|| :  |
Qy     62 PKS-PCEGTECQNGANCVD-QGNRP-VCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQ 118

Db    321 S-D-GTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVY 360
          :    ||: : |:  ||||| | :  ::  || |:|:: |
Qy    119 NWXRXNITLQVFTAEDNGILLYNGGNDHIAVXLYXGHVRFSY 160


RESULT    6
ID   Q91902      PRELIMINARY;      PRT;   721 AA.
AC   Q91902;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   X-DELTA-1.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95319507.
RA   HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT   "Expression of a Delta homologue in prospective neurons in the
RT   chick.";
RL   NATURE 375:787-790(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95319503.
RA   CHITNIS A.B., HENRIQUE D., LEWIS J., ISH-HOROWICZ D., KINTNER C.R.;
RT   "Primary neurogenesis in Xenopus embryos regulated by a homologue of
RT   the Drosophila neurogenic gene Delta.";
RL   NATURE 375:761-766(1995).
DR   EMBL; L42229; G807696; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   721 AA;  79922 MW;  028040EF CRC32;

  Query Match           26.6%;  Score 322;  DB 13;  Length 721;
  Best Local Similarity  46.5%;  Pred. No. 2.37e-39;
  Matches    46;  Conservative   17;  Mismatches  30;  Indels   6;  Gaps   4;

Db    431 CQCQEGFSGRNCDDNLDDCTSFPCQNGGTCQDGINDYSCTCPPGYIGKNCSMP--I-T-K 486
          |:|  |::| || :| |||    ||||: | | :| ||| |: || |  | :|  : : |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    487 --CEHNPCHNGATCHERNNRYVCQCARGYGGNNCQFLLP 523
            ||   |:||| | :: || ||||  |:|| :|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT    7
ID   Q25059      PRELIMINARY;      PRT;   406 AA.
AC   Q25059;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN III (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33862; G499688; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 7.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   406 AA;  43475 MW;  45E6EE2C CRC32;

  Query Match           25.5%;  Score 309;  DB 5;  Length 406;
  Best Local Similarity  44.2%;  Pred. No. 6.34e-37;
  Matches    42;  Conservative   19;  Mismatches  28;  Indels   6;  Gaps   4;

Db     74 CNCIPGFDGDNCENNINECASNPCQNGGVCIDGVNGFVCTCQPGYTGTLCET--DI-D-- 128
          |:|:||: |||| :| ::|  : ||||: |:| ||:: | |  ||:| |||    :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    129 E-CASNPCQNGGVCTDLVNMYTCDCLAGFTGSNCE 162
          | :  ||||: | |  |   |:||:|| |::||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT    8
ID   Q90656      PRELIMINARY;      PRT;   728 AA.
AC   Q90656;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE PROTEIN C-DELTA-1.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 95319507.
RA   HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT   "Expression of a Delta homologue in prospective neurons in the
RT   chick.";
RL   NATURE 375:787-790(1995).
DR   EMBL; U26590; G882412; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   728 AA;  79861 MW;  7439F575 CRC32;
```

```
  Query Match          25.1%;  Score 304;  DB 13;  Length 728;
  Best Local Similarity  43.4%;  Pred. No. 5.40e-36;
  Matches   43;  Conservative   21;  Mismatches  29;  Indels   6;  Gaps   4;

Db    436 CQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCPPGYNGKNCSTP--V-S-R 491
          |:| :|::| :| :| |||    | ||: | | || ||| |: ||:|  |  |  : : :
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    492 --CEHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
            ||   |:||| | :::|| ||:|  |:|| :|: ||:
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLS 102


RESULT   9
ID   Q90285      PRELIMINARY;      PRT;  1476 AA.
AC   Q90285; Q98847;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PUTATIVE EXTRACELLULAR AND CYTOPLASMIC FRAGMENT OF NOTCH-3 HOMOLOG
DE   (FRAGMENT).
OS   CARASSIUS AURATUS (GOLDFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; CYPRININAE; CARASSIUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA (10D POST-OPTIC NERVE CRUSH);
RA   SULLIVAN S.A., BARTHEL L.K., LARGENT B.L., RAYMOND P.A.;
RL   SUBMITTED (APR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U09191; G1617281; -.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PFAM; PF00008; EGF; 11.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   ANK REPEAT; GLYCOPROTEIN.
FT   NON_TER       1      1
FT   NON_TER    1476   1476
SQ   SEQUENCE   1476 AA;  160385 MW;  D6077129 CRC32;

  Query Match          25.0%;  Score 303;  DB 13;  Length 1476;
  Best Local Similarity  44.8%;  Pred. No. 8.28e-36;
  Matches   43;  Conservative   12;  Mismatches  40;  Indels   1;  Gaps   1;

Db    186 CDCMPGYEGDNCEREVNECQSHPCQNGGTCIDLVGHYIRSCPPGTLGVLCEINGDDCATP 245
          |:||||| |||| : ::|: | ||||: |:| |  |   |: |  | ||||     ::|
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHL-PAP 62

Db    246 SWPRGMPKCQNNGTCVDRVGGYRCNCPPGFTGERCE 281
          |     ||| : |||:     | | ||| |  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   10
ID   Q25058      PRELIMINARY;      PRT;   529 AA.
AC   Q25058;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN IA (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33861; G499686; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 10.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   529 AA;  55543 MW;  6385F322 CRC32;

  Query Match          24.8%;  Score 301;  DB 5;  Length 529;
  Best Local Similarity  43.2%;  Pred. No. 1.95e-35;
  Matches   41;  Conservative   19;  Mismatches  29;  Indels   6;  Gaps   2;

Db    273 CSCVQGFTGSDCETNINECASGPCQNGGTCVDGVNGFVCQCPPNYTGTYCEIS--LDA-- 328
          | |: |::| :|  | ::|    ||||: |:| ||:: | |:  |:|  |||:  | |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    329 --CSSMPCQNGATCVNVGANYICECPPGFAGQNCE 361
            | :  ||||| ||: |   :|:| |||:| :||
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   11
ID   P87357      PRELIMINARY;      PRT;   717 AA.
AC   P87357;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAD TRANSMEMBRANE PROTEIN PRECURSOR.
GN   DELTAD.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97346722.
RA   DORNSEIFER P., TAKKE C., CAMPOS-ORTEGA J.A.;
RT   "Overexpression of a zebrafish homologue of the Drosophila neurogenic
RT   gene Delta perturbs differentiation of primary neurons and somite
RT   development.";
RL   MECH. DEV. 63:159-171(1997).
DR   EMBL; Y11760; E307461; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   SIGNAL; TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    717       DELTAD TRANSMEMBRANE PROTEIN.
SQ   SEQUENCE   717 AA;  79061 MW;  5CC32ECA CRC32;

  Query Match          24.6%;  Score 298;  DB 13;  Length 717;
  Best Local Similarity  40.6%;  Pred. No. 7.00e-35;
  Matches   41;  Conservative   22;  Mismatches  32;  Indels   6;  Gaps   5;

Db    427 CQCPEGFTGTHCEDNIDECATYPCQNGGTCQDGLSDYTCTCPPGYTGKNC-TSA-V-N-K 482
          |:|  |::| :| :| |:|    ||||: | | :: |:| |: ||:|  |  :: :   |
Qy      4 CECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPK 63

Db    483 --CLHNPCHNGATCHEMDNRYVCACIPGYGGRNCQFLLPEN 521
            |    |:||| | :  || || |:||:|| :|: ||: |
Qy     64 SPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECEKLLSVN 104


RESULT   12
ID   O88278      PRELIMINARY;      PRT;  3313 AA.
AC   O88278;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF2.
GN   MEGF2.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
```

```
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB011528; D1033422; -.
DR   PROSITE; PS00232; CADHERIN; 7.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   3313 AA;  359348 MW;  5D724643 CRC32;

  Query Match           24.5%;  Score 297;  DB 11;  Length 3313;
  Best Local Similarity  32.7%;  Pred. No. 1.07e-34;
  Matches   54;  Conservative   39;  Mismatches  59;  Indels  13;  Gaps   9;

Db   1411 RCRCPPGFTGDFCETELDLCYSNPCRNGGACARREGGYTCVCRPRFTGEDCELDTE--AG 1468
           || | ||::|| |  : | |  : |:||: |     :|:|:|   ::|: ||:     |
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db   1469 R--CVPGVCRNGGTCTNAPNGGFRCQCPAGGAFEGPRCE-VAARSF-PPSSFVMFRGLRQ 1524
           :  |    |:||: | :  |    ||| :|  | || || : : :|    ::: |  |:
Qy     63 KSPCEGTECQNGANCVDQGNRPV-CQCLPG--FGGPECEKLLSVNFVDRDTYLQFTDLQN 119

Db   1525 RFHLTLSLSFATVQPSGLLFYNGRLNEKHDFLALELVAGQVRLTY 1569
           : :  ::|   | : :|:|:|||  |: |  :|: |  |:||::|
Qy    120 WXRXNITLQVFTAEDNGILLYNGG-ND-H--IAVXLYXGHVRFSY 160


RESULT   13
ID   Q06008      PRELIMINARY;      PRT;  1203 AA.
AC   Q06008;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH PROTEIN HOMOLOG 2 (MOTCH B PROTEIN) (FRAGMENT).
GN   NOTCH2 OR MOTCH B.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
DR   EMBL; X68279; G287990; -.
DR   MGD; MGI:97364; NOTCH2.
DR   PFAM; PF00008; EGF; 27.
DR   PFAM; PF00066; notch; 1.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT.
FT   NON_TER       1      1
FT   NON_TER    1203   1203
SQ   SEQUENCE   1203 AA;  128982 MW;  A5A95551 CRC32;

  Query Match           24.3%;  Score 294;  DB 11;  Length 1203;
  Best Local Similarity  44.8%;  Pred. No. 3.85e-34;
  Matches   43;  Conservative   17;  Mismatches  31;  Indels   5;  Gaps   4;

Db    855 RCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCE--ENI-D- 910
           ||||:||| | ||  : |:|::: ||||: |:| || : | |: |  | |||   ::
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    911 ECA-GGPHCLNGGQCVDRIGGYTCRCLPGFAGERCE 945
           :  |  | ||: |||:     |:|||||:|  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


RESULT   14
ID   Q92566      PRELIMINARY;      PRT;  2408 AA.
AC   Q92566;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELOBLAST KIAA0279 (FRAGMENT).
GN   KIAA0279.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA RES. 3:321-329(1996).
DR   EMBL; D87469; D1014097; -.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PFAM; PF00008; EGF; 6.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF00054; laminin_G; 1.
KW   GLYCOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   2408 AA;  261740 MW;  CDBA2001 CRC32;

  Query Match           24.3%;  Score 295;  DB 4;  Length 2408;
  Best Local Similarity  30.8%;  Pred. No. 2.52e-34;
  Matches   44;  Conservative   40;  Mismatches  49;  Indels  10;  Gaps   9;

Db    758 RCRCPPGFTGDYCETEVDLCYSRPCGPHGRCRSREGGYTCLCRDGYTGEHCEVSAR-SG- 815
           || | ||::|| |  : | |  : |    ::|     :|:||| :||:|: ||:::: ::
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db    816 R--CTPGVCKNGGTCVNLLVGGFKCDCPSGDFEKPYCQ-VTTRSFPAH-SFITFRGLRQR 871
           :  |    | ||: ||:       |:| :| |  | |: : : :|  : ::: |  |: :
Qy     63 KSPCEGTECQNGANCVDQGNRPV-CQCLPG-FGGPECEKLLSVNFVDRDTYLQFTDLQNW 120

Db    872 FHFTLALS-FATKERDGLLLYNG 893
           :  ::|  |:: : :|:|||||
Qy    121 XRXNITLQVFTAED-NGILLYNG 142


RESULT   15
ID   O35516      PRELIMINARY;      PRT;  2470 AA.
AC   O35516;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CELL SURFACE PROTEIN.
GN   NOTCH2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RA   HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL   SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D32210; D1022953; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
```

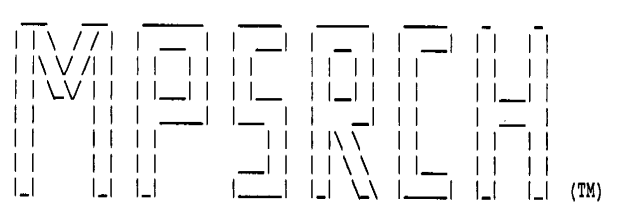
```
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match           24.3%;  Score 294;  DB 11;  Length 2470;
  Best Local Similarity  44.8%;  Pred. No. 3.85e-34;
  Matches   43;  Conservative   17;  Mismatches  31;  Indels   5;  Gaps   4;

Db   1170 RCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCE--ENI-D- 1225
         ||||:||| | ||  : |:|::: |||| |:| || : | |: |  | |||   ::
Qy      3 RCECMPGYAGDNCSENQDDCRDHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAP 62

Db   1226 ECA-GGPHCLNGGQCVDRIGGYTCRCLPGFAGERCE 1260
           :  |  | ||: |||:     |:|||||:|  ||
Qy     63 KSPCEGTECQNGANCVDQGNRPVCQCLPGFGGPECE 98


Search completed: Fri May 28 08:56:06 1999
Job time : 55 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:58:00 1999; MasPar time 6.52 Seconds
335.829 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-6
Description: (1-103) from US09191647.pep
Perfect Score: 813
Sequence: 1 QCHISDQGEPYCLCQPGFSG.....GSSFVEEVERHLECGCLACS 103

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

atistics: Mean 27.967; Variance 116.252; scale 0.241

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	415	51.0	1534	30	W46966	Amino acid sequence o	5.16e-30
2	139	17.1	1480	5	R25079	Drosophila SLIT prote	7.60e-04
3	132	16.2	1872	36	W68510	Partial human Notch-3	3.08e-03
4	132	16.2	2321	36	W49698	Human Notch3 protein.	3.08e-03
5	121	14.9	379	5	R25565	Beta-IG-M1.	2.68e-02
6	118	14.5	1055	29	W44298	Human serrate 2 prote	4.81e-02
7	118	14.5	1212	29	W44299	Human serrate 2.	4.81e-02
8	118	14.5	1257	19	W05834	Human Serrate-2 (HJ2)	4.81e-02
9	116	14.3	381	26	W35730	Human cysteine rich p	7.09e-02
10	116	14.3	381	26	W35957	Human monocyte mature	7.09e-02
11	108	13.3	179	37	W75100	Human secreted protei	3.28e-01
12	107	13.2	375	16	R90919	Connective tissue gro	3.97e-01
13	105	12.9	193	3	P60463	Sequence of C-terminu	5.79e-01
14	105	12.9	2050	38	W73499	Von Willebrand factor	5.79e-01
15	105	12.9	2813	3	P60462	Sequence of human von	5.79e-01
16	103	12.7	612	28	W39256	Human partial mature	8.43e-01
17	103	12.7	737	28	W39257	Human membrane protei	8.43e-01
18	102	12.5	2813	3	P60053	Sequence of von Wille	1.02e+00
19	101	12.4	685	37	W80813	Nucleotide sequence o	1.23e+00
20	100	12.3	157	21	W11730	H-Delta-1 polypeptide	1.48e+00
21	100	12.3	520	25	W18348	Proliferation and dif	1.48e+00
22	100	12.3	660	21	W11725	H-Delta-1 polypeptide	1.48e+00
23	100	12.3	702	25	W18349	Proliferation and dif	1.48e+00
24	100	12.3	723	25	W18353	Proliferation and dif	1.48e+00
25	99	12.2	727	21	W11719	C-Delta-1 polypeptide	1.78e+00
26	99	12.2	740	21	W00876	C-Delta-1 polypeptide	1.78e+00
27	99	12.2	1036	25	W18351	Proliferation and dif	1.78e+00
28	99	12.2	1187	25	W18352	Proliferation and dif	1.78e+00
29	99	12.2	1208	28	W40827	Human Jagged protein.	1.78e+00
30	99	12.2	1218	19	W05833	Human Serrate-1 (HJ1)	1.78e+00
31	99	12.2	1218	29	W44301	Human serrate 1.	1.78e+00
32	99	12.2	1218	25	W18354	Proliferation and dif	1.78e+00
33	98	12.1	32	20	W12907	Modified ligand speci	2.14e+00
34	98	12.1	374	27	W37497	Human TMP-2.	2.14e+00
35	98	12.1	374	22	W07663	Human transforming gr	2.14e+00
36	98	12.1	1193	19	W05835	Chick Serrate.	2.14e+00
37	95	11.7	385	10	R56167	Neuroendocrine tumor	3.72e+00
38	95	11.7	833	6	R28960	Delta D11.	3.72e+00
39	94	11.6	33	20	W12909	Modified ligand speci	4.47e+00
40	94	11.6	208	16	R92897	Human HBEGF precursor	4.47e+00
41	94	11.6	208	13	R66190	Diphtheria toxin (DT)	4.47e+00
42	94	11.6	208	16	R92898	Monkey HBEGF precurso	4.47e+00
43	94	11.6	208	16	R80787	Monkey precursor hepa	4.47e+00
44	94	11.6	333	16	R92900	HBEGF Val Met saporin	4.47e+00
45	94	11.6	334	16	R92921	Saporin-HBEGF protein	4.47e+00

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 51.0%; Score 415; DB 30; Length 1534;
Best Local Similarity 49.0%; Pred. No. 5.16e-30;
Matches 51; Conservative 22; Mismatches 30; Indels 1; Gaps 1;

Db 1431 hcqasgtkgahcvcdpgfsgelceqesecrgdpvrdfhqvqrgyaicqttrplswvecrg 1490
Qy 1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRG 60

Db 1491 scpgqgccqglrlkrrkftfecsdgtsfaeevekptkcgcalca 1534

```
Qy     61 GC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 103


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match          17.1%;  Score 139;  DB 5;  Length 1480;
  Best Local Similarity  35.7%;  Pred. No. 7.60e-04;
  Matches   15;  Conservative   9;  Mismatches  18;  Indels   0;  Gaps   0;

Db   1434 cvggcgnqccaakivrrrkvrmvcsnnrkyiknldivrkcgc 1475
         | |||| ||| :  :|||   |::  :: :::   |||
Qy     58 CRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    3
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 328
FT                   /note= "encoded by NAN"
FT   Misc_difference 401
FT                   /note= "encoded by GNN"
FT   Misc_difference 403
FT                   /note= "encoded by GNC"
FT   Misc_difference 406
FT                   /note= "encoded by GNN"
FT   Misc_difference 409
FT                   /note= "encoded by NNT"
FT   Misc_difference 420
FT                   /note= "encoded by GNC"
FT   Misc_difference 706
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
SQ   Sequence   1872 AA;

  Query Match          16.2%;  Score 132;  DB 36;  Length 1872;
  Best Local Similarity  45.5%;  Pred. No. 3.08e-03;
  Matches   15;  Conservative   9;  Mismatches   7;  Indels   2;  Gaps   2;
```

```
Db    994 qc-vdedsshycvcpegrtgshceqevdpclaq 1025
          || : :::  ||:|  | :| ||:|| :|||:|
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32
```

RESULT 4
ID W49698 standard; Protein; 2321 AA.
AC W49698;
DT 21-DEC-1998 (first entry)
DE Human Notch3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leukoencephalopathy; therapy.
OS Homo sapiens.
PN FR2751986-A1.
PD 06-FEB-1998.
[redacted] 16-APR-1997; 004680.
01-AUG-1996; FR-009733.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR WPI; 98-133138/13.
DR N-PSDB; V57001.
PT Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leukoencephalopathy
PS Claim 2; Fig 1.1-1.8; 45pp; French.
CC This sequence represents the human Notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of
CC the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC and leukoencephalopathy (CADASIL) type. Blocking expression of a
CC mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC gene or protein can be used to treat CADASIL or related disorders.
SQ Sequence 2321 AA;

```
  Query Match          16.2%;  Score 132;  DB 36;  Length 2321;
  Best Local Similarity 45.5%;  Pred. No. 3.08e-03;
  Matches   15;  Conservative   9;  Mismatches   7;  Indels   2;  Gaps   2;

Db   1060 qc-vdedsshycvcpegrtgshceqevdpclaq 1091
          || : :::  ||:|  | :| ||:|| :|||:|
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32
```

[redacted]ULT 5
ID R25565 standard; Protein; 379 AA.
AC R25565;
DT 18-JAN-1993 (first entry)
DE Beta-IG-M1.
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW embryo; fibroblasts; TGF-beta.
OS Mus musculus.
PN EP-495674-A.
PD 22-JUL-1992.
PF 17-JAN-1992; 300429.
PR 18-JAN-1991; US-642991.
PR 10-JAN-1992; US-816270.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR WPI; 92-243508/30.
DR N-PSDB; Q26421.
PT TGF-beta induced gene family - encodes proteins involved in
PT growth and differentiation effects of TGF-beta-1
PS Claim 2; Fig 1; 35pp; English.
CC The protein sequence was deduced from the DNA sequence obtd. by
CC screening a cDNA library made from AKR-2B mouse cells induced with
CC TGF-beta1 and cyclohexamide with two probes from untreated AKR-2B
CC mRNA and AKR-2B mRNA from cells treated with cyclohexamide and TGF-
CC beta1. The proteins encoded by hybridising colonies (beta-IG-M1 and
CC beta-IG-M2) contain 38 Cys residues and are induced by TGF-beta1.
CC Beta-IG-M1 displays 80 percent homology to the CEF-10 protein
CC induced by v-src in chicken embryo fibroblasts and is identical
CC to the protein encoded by cyr61, an immediate early response gene
CC induced in quiescent BALB 3T3 cells by serum treatment. Residues
CC 49-56 of beta-IG-M1 conform to the GCGCCXXC motif reported in the
CC amino half of insulin-like growth factor (IGF) binding proteins.
CC The C-terminal Cys rich region of beta-IG-M1, -M2 and CEF-10 contain
CC an amino acid sequence with strong homology to a motif found near the
CC C-terminal of the malarial circumsporozoite (CS) protein, which is
CC highly conserved among all species of malarial parasites sequenced
CC to date (designated region II). This motif is also found in
CC other proteins which have cell adhesive properties that mediate
CC cell-cell and cell-extracellular matrix interactions, such as
CC properdin, thrombospondin, and TRAP. The proteins encoded by
CC TGF-beta induced genes are likely to be involved in mediation of
CC the biological effects of TGF-beta relating to cell growth and
CC differentiation. See also R25566.
SQ Sequence 379 AA;

```
  Query Match          14.9%;  Score 121;  DB 5;  Length 379;
  Best Local Similarity 33.3%;  Pred. No. 2.68e-02;
  Matches   19;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db   298 yagcssvkkyrpkyc-gscvdgrcctplqtrtvkmrfrcedgemfsknvmmiqsckc 353
          ||:|:: |  : | |:|  :|| | ::: |  |:| || |  :|    | |
Qy    44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99
```

RESULT 6
ID W44298 standard; Protein; 1055 AA.
AC W44298;
DT 19-JUN-1998 (first entry)
DE Human serrate 2 protein fragment.
KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.
PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15181.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Claim 2; Page 57-62; 103pp; Japanese.
CC The present sequence represents a human serrate 2 protein fragment. The
CC present invention also describes a method for the preparation of the
CC polypeptides, and antibodies binding to the polypeptide and its
CC fragments. The polypeptide and its fragments expressed by the serrate-2
CC gene can be used to inhibit stem (especially blood stem) cell
CC differentiation and to inhibit endothelial cell growth. They may be
CC incorporated in a cell culture media for culturing undifferentiated
CC stem cells. They can also be used for treatment of neoplasms such as
CC leukaemia. The antibodies can be used for the diagnosis of malignant
CC tumours.
SQ Sequence 1055 AA;

```
  Query Match          14.5%;  Score 118;  DB 29;  Length 1055;
  Best Local Similarity 36.8%;  Pred. No. 4.81e-02;
  Matches   14;  Conservative   13;  Mismatches   8;  Indels   3;  Gaps   3;

Db   585 rc-vsqpggnfscicdsgftgtycheniddclgqpcrn 621
          :| :|: |, : |:|::||:|  |::: : ||||  |:
Qy     1 QCHISDQGEPY-CLCQPGFSGEHCQQE-NPCLGQVVRE 36
```

RESULT 7
ID W44299 standard; Protein; 1212 AA.
AC W44299;
DT 19-JUN-1998 (first entry)
DE Human serrate 2.

KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.
PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15181.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Claim 3; Page 62-68; 103pp; Japanese.
CC The present sequence represents human serrate 2. The present invention
CC also describes a method for the preparation of the polypeptides, and
CC antibodies binding to the polypeptide and its fragments. The polypeptide
CC and its fragments expressed by the serrate-2 gene can be used to inhibit
CC stem (especially blood stem) cell differentiation and to inhibit
CC endothelial cell growth. They may be incorporated in a cell culture
CC media for culturing undifferentiated stem cells. They can also be used
CC for treatment of neoplasms such as leukaemia. The antibodies can be used
CC for the diagnosis of malignant tumours.
SQ Sequence 1212 AA;

Query Match 14.5%; Score 118; DB 29; Length 1212;
Best Local Similarity 36.8%; Pred. No. 4.81e-02;
Matches 14; Conservative 13; Mismatches 8; Indels 3; Gaps 3;

Db 585 rc-vsqpggnfscicdsgftgtycheniddclgqpcrn 621
:| :|: | : |:|::||:| |::: : |||| |:
Qy 1 QCHISDQGEPY-CLCQPGFSGEHCQQE-NPCLGQVVRE 36

RESULT 8
ID W05834 standard; Protein; 1257 AA.
AC W05834;
DT 28-JAN-1997 (first entry)
DE Human Serrate-2 (HJ2).
KW Serrate-2; human jagged-2; HJ2; Notch; cell differentiation;
KW cell fate; central nervous system; cancer; tissue repair; therapy;
KW diagnosis; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..912
FT /label= Extracellular_domain
FT /note= "a deletion in the encoding cDNA clone
FT results in loss of part of the Serrate-2
FT signal peptide and beginning of the DSL
FT domain"
FT domain 26..70
FT /label= DSL
FT /note= "region of homology with Drosophila Delta
FT and Serrate, predicted to mediate binding
FT with Notch"
FT domain 75..735
FT /label= ELR
FT /note= "epidermal growth factor-like repeat domain"
FT region 75..105
FT /label= ELR1
FT region 106..140
FT /label= ELR2
FT region 141..180
FT /label= ELR3
FT region 181..218
FT /label= ELR4
FT region 219..256
FT /label= ELR5
FT region 257..294
FT /label= ELR6
FT region 295..331
FT /label= ELR7
FT region 332..369
FT /label= ELR8
FT region 370..407
FT /label= ELR9
FT region 408..435
FT /label= Partial_ELR
FT region 436..469
FT /label= Partial_ELR
FT region 470..507
FT /label= ELR10
FT region 508..545
FT /label= ELR11
FT region 546..584
FT /label= ELR12
FT region 585..622
FT /label= ELR13
FT region 623..660
FT /label= ELR14
FT region 664..701
FT /label= ELR15
FT region 702..718
FT /label= Partial_ELR
FT region 719..735
FT /label= Partial_ELR
FT domain 913..933
FT /label= Transmembrane_domain
FT domain 934..1257
FT /label= Intracellular_domain
PN WO9627610-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03172.
PR 07-MAR-1995; US-400159.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI Lewis JH, Mann RS, Myat AM;
DR WPI; 96-425379/42.
DR N-PSDB; W05834.
PT Vertebrate Serrate protein and related DNA - used to treat or
PT prevent malignancies characterised by increased Notch activity.
PS Claim 5; Page 104-107; 161pp; English.
CC Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC for the zygotic neurogenic locus Notch, and are believed to play a
CC major role in determining cell fates (differentiation) in the
CC central nervous system. Their amino acid sequences were deduced
CC from cDNA clones (see also T40090-91) isolated from human foetal
CC brain cDNA libraries. The proteins, antibodies raised to them,
CC and encoding nucleic acids can be used in the detection of
CC Serrate sequences and in the treatment of disorders of cell fate
CC or differentiation, partic. cancer, nervous system disorders
CC and in tissue repair or regeneration.
SQ Sequence 1257 AA;

Query Match 14.5%; Score 118; DB 19; Length 1257;
Best Local Similarity 36.8%; Pred. No. 4.81e-02;
Matches 14; Conservative 13; Mismatches 8; Indels 3; Gaps 3;

Db 441 rc-vsqpggnfscicdsgftgtycheniddclgqpcrn 477
:| :|: | : |:|::||:| |::: : |||| |:
Qy 1 QCHISDQGEPY-CLCQPGFSGEHCQQE-NPCLGQVVRE 36

RESULT 9
ID W35730 standard; Protein; 381 AA.
AC W35730;
DT 27-MAR-1998 (first entry)
DE Human cysteine rich protein 61 (Cyr61).
KW Cysteine rich protein 61; Cyr61; human;
KW extracellular matrix signalling molecule; cell adhesion;
KW cell migration; cell proliferation; angiogenesis; chondrogenesis;
KW oncogenesis; haematostasis; wound healing; organ regeneration.
OS Homo sapiens.
PN WO9733995-A2.

PD 18-SEP-1997.
PF 14-MAR-1997; U04193.
PR 15-MAR-1996; US-013958.
PA (MUNI-) MUNIN CORP.
PI Lau LF;
DR WPI; 97-470875/43.
DR N-PSDB; T94699.
PT Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT modulate e.g. haematostasis, induce wound healing, promote organ
PT regeneration etc
PS Claim 2; Page 112-113; 133pp; English.
CC This protein sequence comprises human cysteine rich protein 61
CC (Cyr61), an extracellular matrix signalling molecule. Its amino
CC acid sequence was deduced from a human placental cDNA clone (see
CC T94699). Cyr61 polypeptides can be expressed in transformed or
CC transfected host cells. Cyr61 can be used to modulate
CC haematostasis, induce wound healing in a tissue, promote organ
CC regeneration, improve tissue grafting or promote bone or prothesis
CC implantation (claimed). It can also be used to screen for a
CC modulator of angiogenesis, chondrogenesis, oncogenesis, cell
CC adhesion, cell migration, cell proliferation, expand a population
CC of undifferentiated haematopoietic stem cells in culture and to
CC screen for a mitogen (claimed). Ex vivo methods for using
CC mammalian extracellular matrix signalling molecules to prepare
CC blood products are also provided.
SQ Sequence 381 AA;

Query Match 14.3%; Score 116; DB 26; Length 381;
Best Local Similarity 33.3%; Pred. No. 7.09e-02;
Matches 19; Conservative 10; Mismatches 26; Indels 2; Gaps 2;

Db 300 yagclsvkkyrpkyc-gscvdgrcctpqltrtvkmrfrcedgetfsknvmmiqsckc 355
Qy 44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99

RESULT 10
ID W35957 standard; Protein; 381 AA.
AC W35957;
DT 05-MAR-1998 (first entry)
DE Human monocyte mature differentiation factor.
KW Human; monocyte; mature; differentiation factor; MMDF; macrophage;
KW cancer; immune activator; tissue culture; infectious disease.
OS Homo sapiens.
PN J09234079-A.
09-SEP-1997.
04-MAR-1996; 075236.
PR 04-MAR-1996; JP-075236.
PA (TOYM) TOYOBO KK.
DR WPI; 97-497320/46.
DR N-PSDB; T97142.
PT A monocyte mature differentiation factor - useful for the long term
PT tissue culture of macrophage(s)
PS Claim 9; Page 12-13; 22pp; Japanese.
CC The present sequence represents a monocyte mature differentiation
CC factor (MMDF) which maintains the life of macrophages for long periods
CC in liquid culture. MMDF can be used as an anti-cancer agent, an immune
CC activator and to treat infectious diseases.
SQ Sequence 381 AA;

Query Match 14.3%; Score 116; DB 26; Length 381;
Best Local Similarity 33.3%; Pred. No. 7.09e-02;
Matches 19; Conservative 10; Mismatches 26; Indels 2; Gaps 2;

Db 300 yagclsvkkyrpkyc-gscvdgrcctpqltrtvkmrfrcedgetfsknvmmiqsckc 355
Qy 44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99

RESULT 11
ID W75100 standard; Protein; 179 AA.
AC W75100;
DT 28-JAN-1999 (first entry)
DE Human secreted protein encoded by gene 44 clone HE8CJ26.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 179
FT /label= unknown
PN WO9839446-A2.
PD 11-SEP-1998.
PF 06-MAR-1998; U04492.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043315.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.

PR 06-JUN-1997; US-048974.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057761.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-609887/51.
DR N-PSDB; V34197.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 305; 447pp; English.
CC This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line.
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V34145) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 70 novel genes and their fragments (nucleic acid
CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 70 polynucleotides, based on
CC which tissues they are most highly expressed in (see V34154 for described
CC uses).
SQ Sequence 179 AA;

Query Match 13.3%; Score 108; DB 37; Length 179;
Best Local Similarity 40.6%; Pred. No. 3.28e-01;
Matches 13; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 121 chmlsr-dtyectcqvgftgkecqwtdaclsh 151
|| : : | | || ||:| || ::||::
Qy 2 CHISDQGEPY-CLCQPGFSGEHCQQENPCLGQ 32

RESULT 12
ID R90919 standard; Protein; 375 AA.
AC R90919;
DT 25-JUN-1996 (first entry)
DE Connective tissue growth factor-2.
KW CTGF-2; connective tissue growth factor-2; secreted protein;
KW cartilagenous growth; skeletal; embryo; cell growth; morphogenesis;
KW insulin-like growth factor; fibroblast growth factor; Cry61.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..24
FT /label= signal_peptide
FT protein 25..375
FT /label= mature_protein
PN WO9601896-A.
PD 25-JAN-1996.
PF 12-JUL-1994; U07736.
PR 12-JUL-1994; WO-U07736.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Adams MD, Li H;
DR WPI; 96-097626/10.
DR N-PSDB; T12653.
PT Connective tissue growth factor-2 and DNA encoding it - useful to
PT enhance the repair of connective and support tissue, and to enhance
PT wound healing
PS Claim 1; Fig 1A-C; 46pp; English.
CC Connective tissue growth factor-2 (CTGF-2) is encoded by a cDNA
CC (T12653) isolated from a human foetal lung cDNA library. The GTGF
CC polypeptides are structurally and functionally related to a family
CC of growth factors which include IGF (Insulin-like growth factor),
CC PDGF (platelet-derived growth factor), and FGF (fibroblast growth
CC factor). CTGF-2 exhibits 89 percent identity and 93 percent similarity
CC to Cry61. Cry61 is a growth factor-inducible immediate early gene
CC initially identified in serum-stimulated mouse fibroblasts. It encodes
CC a member of an emerging family of secreted proteins which are also a
CC group of cysteine-rich proteins. This group of GFs are important for
CC normal growth, differentiation, morphogenesis of the cartilaginous
CC skeleton of an embryo and cell growth.
SQ Sequence 375 AA;

Query Match 13.2%; Score 107; DB 16; Length 375;
Best Local Similarity 34.7%; Pred. No. 3.97e-01;
Matches 17; Conservative 9; Mismatches 21; Indels 2; Gaps 2;

Db 301 yagclsvkkyrpkyc-gscvdgrcctpqltrtvkmrfpcedgetfsknv 348
||:| : | : | |:| :|| | :: | | | || :| :|
Qy 44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEV 91

RESULT 13
ID P60463 standard; Protein; 193 AA.
AC P60463;
DT 25-JUN-1991 (first entry)
DE Sequence of C-terminus of von Willebrand Factor (VWF).
KW Chronic renal failure; therapy; factor VIII C.
OS Homo sapiens.
FH Key Location/Qualifiers
FT modified_site 15
FT /label= Potential N-glycosylation site
FT modified_site 170
FT /label= Potential N-glycosylation site
PN WO8606096-A.
PD 23-OCT-1986.
PF 10-APR-1986; U00760.
PR 11-APR-1985; US-722108.
PA (CHIL-) CHILDRENS MED CENT.
PA (GINS/) GINSBURG D.
PI Ginsburg D, Orkin SH, Kaufman RJ;
DR WPI; 86-291663/44.
DR N-PSDB; N60405.
PT Pure Von Willebrand Factor - produced using an expression vector
PT including a DNA sequence encoding functional VWF protein
PS Disclosure; Table 1, Page 9; 54pp; English.
CC cDNA clones pVWH33, pVWH5 and PVWE6 which span 9 kb pairs of DNA and

CC encompass the entire protein coding region of VWF, were selected to
CC construct full length cDNA (N60404). The pure VWF produced is useful
CC in the treatment of von Willebrand's disease (VWD) and the patients
CC with chronic renal failure whose abnormal bleeding times are
CC corrected by crude cryoprecipitate. Pure VWF can also be used to
CC carry, stabilise and improve the therapeutic efficacy of factor
CC VIII:C.
SQ Sequence 193 AA;

Query Match 12.9%; Score 105; DB 3; Length 193;
Best Local Similarity 41.2%; Pred. No. 5.79e-01;
Matches 14; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

Db 153 ccsptrtepmqvalhctngsvvyhevlnameckc 186
|| ||| : ::||:|| || :|| |
Qy 66 CCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99

[redacted]ULT 14
ID W73499 standard; Protein; 2050 AA.
AC W73499;
DT 26-FEB-1999 (first entry)
DE Von Willebrand factor.
KW Von Willebrand factor; GP1b binding domain; inhibitor; re-occlusion;
KW platelet aggregation; cerebrovascular disorder; cardiovascular disorder;
KW angioplasty; thrombi-containing platelet-rich aggregate; thrombosis;
KW therapy.
OS Homo sapiens.
PN US5849536-A.
PD 15-DEC-1998.
PF 30-NOV-1994; 347594.
PR 01-MAR-1991; WO-U01416.
PR 02-MAR-1990; US-487767.
PR 03-SEP-1991; US-753815.
PR 22-JUN-1993; US-080690.
PR 30-NOV-1994; US-347594.
PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.
PI Garfinkel L, Richter T;
DR WPI; 99-069733/06.
DR N-PSDB; V08901.
PT Polypeptide comprising von Willebrand factor GP1b binding domain -
PT useful as platelet aggregation inhibitor
PS Disclosure; Fig 12; 62pp; English.
CC This sequence represents the mature human Von Willebrand factor protein.
CC GP1b binding domain of this sequence represents the protein of the
CC invention. The protein is used for inhibiting platelet aggregation,
CC especially for treating cerebrovascular disorders. It is also used for
CC treating cardiovascular disorders, especially acute myocardial infarction
CC or angina. The protein is also used for inhibiting platelet aggregation
CC before, during or after angioplasty, thrombolytic treatment or coronary
CC bypass surgery, for maintaining blood vessel potency before, during or
CC after coronary bypass surgery. It can also be used for inhibiting
CC thrombosis, optionally associated with an inflammatory response, for
CC inhibiting platelet adhesion to damaged vascular surfaces, for preventing
CC platelet adhesion to prosthetic materials or devices, for inhibiting
CC re-occlusion after angioplasty or thrombolysis, or for thrombolytic
CC treatment of thrombi-containing platelet-rich aggregates.
SQ Sequence 2050 AA;

Query Match 12.9%; Score 105; DB 38; Length 2050;
Best Local Similarity 41.2%; Pred. No. 5.79e-01;
Matches 14; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

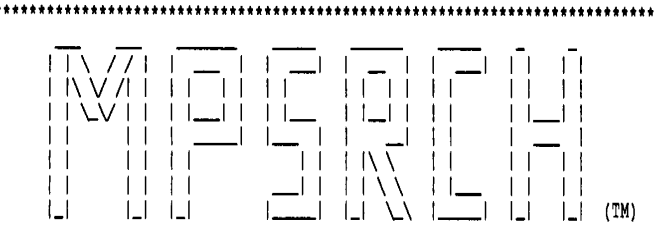
Db 2010 ccsptrtepmqvalhctngsvvyhevlnameckc 2043
|| ||| : ::||:|| || :|| |
Qy 66 CCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99

RESULT 15
ID P60462 standard; Protein; 2813 AA.
AC P60462;
DT 25-JUN-1991 (first entry)
DE Sequence of human von Willebrand Factor (VWF) precursor.
KW Chronic renal failure; therapy; factor VIII C.
OS Homo sapiens.
PN WO8606096-A.
PD 23-OCT-1986.
PF 10-APR-1986; U00760.
PR 11-APR-1985; US-722108.
PA (CHIL-) CHILDRENS MED CENT.
PA (GINS/) GINSBURG D.
PI Ginsburg D, Orkin SH, Kaufman RJ;
DR WPI; 86-291663/44.
DR N-PSDB; N60404.
PT Pure Von Willebrand Factor - produced using an expression vector
PT including a DNA sequence encoding functional VWF protein
PS Disclosure; Table 2, Pages 18-36A; 54pp; English.
CC cDNA clones pVWH33, pVWH5 and PVWE6 which span 9 kb pairs of DNA and
CC encompass the entire protein coding region of VWF, were selected to
CC construct full length cDNA (N60404). The pure VWF produced is useful
CC in the treatment of von Willebrand's disease (VWD) and the patients
CC with chronic renal failure whose abnormal bleeding times are
CC corrected by crude cryoprecipitate. Pure VWF can also be used to
CC carry, stabilise and improve the therapeutic efficacy of factor
CC VIII:C.
SQ Sequence 2813 AA;

Query Match 12.9%; Score 105; DB 3; Length 2813;
Best Local Similarity 41.2%; Pred. No. 5.79e-01;
Matches 14; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

Db 2773 ccsptrtepmqvalhctngsvvyhevlnameckc 2806
|| ||| : ::||:|| || :|| |
Qy 66 CCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99

Search completed: Fri May 28 08:58:51 1999
Job time : 51 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:59:09 1999; MasPar time 7.08 Seconds
582.525 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-6
Description: (1-103) from US09191647.pep
Perfect Score: 813
Sequence: 1 QCHISDQGEPYCLCQPGFSG.....GSSFVEEVERHLECGCLACS 103

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 36.826; Variance 62.401; scale 0.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	150	18.5	1469	2	B36665	slit protein 2 precur	5.93e-13
2	139	17.1	1480	2	A36665	slit protein 1 precur	8.26e-11
3	132	16.2	2321	2	S78549	notch3 protein - huma	1.79e-09
4	129	15.9	375	2	A41428	CEF-10 protein precur	6.59e-09
5	121	14.9	379	2	A35669	gene CYR61 protein pr	2.01e-07
6	116	14.3	5147	1	IJFFTM	cadherin-related tumo	1.63e-06
7	113	13.9	293	2	B26637	neurogenic repetitive	5.62e-06
8	113	13.9	2139	2	A35672	crumbs protein - frui	5.62e-06
9	113	13.9	2437	2	S42612	transmembrane protein	5.62e-06
10	112	13.8	2318	2	S45306	notch 3 protein - mou	8.46e-06
11	111	13.7	2531	2	S18188	notch protein homolog	1.27e-05
12	111	13.7	2531	2	A46019	gene Notch-1 protein	1.27e-05
13	110	13.5	706	2	H71707	hypothetical protein	1.91e-05
14	110	13.5	2471	2	A49128	cell-fate determining	1.91e-05
15	108	13.3	1203	2	A49175	Motch B protein - mou	4.29e-05
16	108	13.3	2524	2	A35844	Xotch protein - Afric	4.29e-05
17	107	13.2	530	2	A31640	epidermal growth fact	6.41e-05
18	105	12.9	2813	1	VWHU	von Willebrand factor	1.42e-04
19	105	12.9	3707	2	S18252	heparan sulfate prote	1.42e-04
20	105	12.9	4543	1	A53102	alpha-2-macroglobulin	1.42e-04
21	103	12.7	638	2	S08042	proteoglycan core pro	3.14e-04
22	103	12.7	2703	2	A24420	notch protein - fruit	3.14e-04
23	102	12.5	387	2	B49175	Motch A protein - mou	4.65e-04
24	102	12.5	570	2	A48836	fibropellin C precurs	4.65e-04
25	102	12.5	603	2	S28941	coagulation factor XI	4.65e-04
26	102	12.5	615	1	KFHU12	coagulation factor XI	4.65e-04
27	102	12.5	861	2	A48825	Notch homolog Motch p	4.65e-04
28	102	12.5	2555	2	A40043	notch protein homolog	4.65e-04
29	101	12.4	1042	2	A57534	mucin (clone L31) - h	6.87e-04
30	101	12.4	1056	2	A53767	tracheobronchial muci	6.87e-04
31	101	12.4	4391	2	A38096	perlecan precursor -	6.87e-04
32	99	12.2	427	2	JC4915	ags protein precursor	1.49e-03
33	99	12.2	728	2	I50719	C-Delta-1 - chicken	1.49e-03
34	99	12.2	1220	2	A56136	jagged protein precur	1.49e-03
35	96	11.8	383	2	B45484	delta-like dlk homeot	4.70e-03
36	96	11.8	383	2	S53716	homeotic protein dlk	4.70e-03
37	96	11.8	463	2	A36479	milk fat globule memb	4.70e-03
38	96	11.8	3712	2	S18253	laminin alpha-1 chain	4.70e-03
39	95	11.7	385	2	S53718	homeotic protein dlk	6.87e-03
40	95	11.7	385	2	A54785	preadipocyte factor 1	6.87e-03
41	95	11.7	832	2	A31246	neurogenic protein De	6.87e-03
42	95	11.7	833	2	S19087	gene Delta protein pr	6.87e-03
43	95	11.7	880	2	S00670	gene Delta protein pr	6.87e-03
44	95	11.7	1064	2	A40136	fibropellin Ia - sea	6.87e-03
45	95	11.7	3566	2	A40701	tenascin-X precursor	6.87e-03

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
  371-394         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494         #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537         #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695         #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733         #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  846-890         #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061       #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           18.5%;  Score 150;  DB 2;  Length 1469;
  Best Local Similarity  29.9%;  Pred. No. 5.93e-13;
  Matches   29;  Conservative   18;  Mismatches  44;  Indels   6;  Gaps   6;

Db   1372 SNARDGYQCKCKHGQRGRYCDQAASTCRKEQVREYYT-ENDCRSRQPL-KYA-K-CVGGC 1427
         |:  : | | |  |  |  |:|  : |  : |||    ::   |     | : : | |||
Qy      5 SDQGEPY-CLCQPGFSGEHCQQENP-CLGQVVREVIRRQKGYASCATASKVPIMECRGGC 62

Db   1428 GNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGC 1464
          | ||| :   :|||    |::   :: :::    |||
Qy     63 GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing

FEATURE
  66-91         #domain proteoglycan amino-terminal homology #label
                  PAH1\
  101-124       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
  125-148       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
  149-172       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
  173-196       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
  197-220       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
  228-272       #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
  288-313       #domain proteoglycan amino-terminal homology #label
                  PAH2\
  323-346       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
  347-370       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
  371-394       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494       #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537       #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695       #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733       #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  791-814       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
  815-838       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
  846-890       #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061     #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           17.1%;  Score 139;  DB 2;  Length 1480;
  Best Local Similarity  35.7%;  Pred. No. 8.26e-11;
  Matches   15;  Conservative    9;  Mismatches  18;  Indels   0;  Gaps   0;

Db   1434 CVGGCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGC 1475
          | |||| ||| :   :|||    |::   :: :::    |||
Qy     58 CRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    3
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
```

```
   #authors      Joutel, A.; Tournier-Lasserve, E.
   #submission   submitted to the EMBL Data Library, April 1997
   #accession    S78549
     ##molecule_type mRNA
     ##residues      1-2321 ##label JOU1
     ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE        S71825
   #authors      Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
   #journal      Nature (1996) 383:707-710
   #title        Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
   #cross-references MUID:97032728
   #accession    S71825
     ##status       nucleic acid sequence not shown
     ##molecule_type DNA
     ##residues      67-113;138-194;268-333,'G',335-346;536-613;716-765;
                      1240-1279;1815-1888 ##label JOU2
     ##cross-references EMBL:U97669
GENETICS
   #gene         notch3
   #map_position 19p13.1
FUNCTION
   #description  may be involved in pathogenesis of CADASIL, causing a type of
                  stroke and dementia
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS         tandem repeat; transmembrane protein
FEATURE
   318-349           #domain EGF homology #label EGF\
   1838-1870         #domain ankyrin repeat homology #label AN1\
   1871-1903         #domain ankyrin repeat homology #label AN2\
   1905-1937         #domain ankyrin repeat homology #label AN3\
   1938-1970         #domain ankyrin repeat homology #label AN4\
   1971-2003         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2321  #molecular-weight 243657  #checksum 3337

  Query Match          16.2%;  Score 132;  DB 2;  Length 2321;
  Best Local Similarity  45.5%;  Pred. No. 1.79e-09;
  Matches   15;  Conservative    9;  Mismatches   7;  Indels   2;  Gaps   2;

Db   1060 QC-VDEDSSHYCVCPEGRTGSHCEQEVDPCLAQ 1091
          || : :::  ||:|  | :| ||:|| :|||:|
        1 QCHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32


RESULT    4
ENTRY            A41428     #type complete
TITLE            CEF-10 protein precursor - chicken
ORGANISM         #formal_name Gallus gallus #common_name chicken
DATE             03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
                  10-Sep-1997
ACCESSIONS       A41428
REFERENCE        A41428
   #authors      Simmons, D.L.; Levy, D.B.; Yannoni, Y.; Erikson, R.L.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1989) 86:1178-1182
   #title        Identification of a phorbol ester-repressible v-src-inducible
                  gene.
   #cross-references MUID:89145206
   #accession    A41428
     ##status       preliminary
     ##molecule_type mRNA
     ##residues      1-375 ##label SIM
     ##cross-references GB:J04496; NID:g211435; PID:g211436
SUMMARY          #length 375  #molecular-weight 40651  #checksum 1417

  Query Match          15.9%;  Score 129;  DB 2;  Length 375;
  Best Local Similarity  33.3%;  Pred. No. 6.59e-09;
  Matches   19;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db    295 YAGCSSVKKYRPKYC-GSCVDGRCCTPQQTRTVKIRFRCDDGETFTKSVMMIQSCRC 350
          ||:|:: |   : | |:|   :|| | ::: |  |:| || :|   |     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    5
ENTRY            A35669     #type complete
TITLE            gene CYR61 protein precursor - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  16-Dec-1998
ACCESSIONS       A35669; I48319; S16446
REFERENCE        A35669
   #authors      O'Brien, T.P.; Yang, G.P.; Sanders, L.; Lau, L.F.
   #journal      Mol. Cell. Biol. (1990) 10:3569-3577
   #title        Expression of cyr61, a growth factor-inducible
                  immediate-early gene.
   #cross-references MUID:90287146
   #accession    A35669
     ##status       preliminary
     ##molecule_type mRNA
     ##residues      1-379 ##label OAB
     ##cross-references GB:M32490; NID:g192909; PID:g309206
     ##note         the authors translated the codon GAT for residue 337 as
                      Gln
REFERENCE        I48319
   #authors      Latinkic, B.V.; O'Brien, T.P.; Lau, L.F.
   #journal      Nucleic Acids Res. (1991) 19:3261-3267
   #title        Promoter function and structure of the growth
                  factor-inducible immediate early gene cyr61.
   #cross-references MUID:91288203
   #accession    I48319
     ##status       translated from GB/EMBL/DDBJ
     ##molecule_type DNA
     ##residues      1-379 ##label RES
     ##cross-references EMBL:X56790; NID:g50632; PID:g50633
     ##note         the authors did not translate the codon for residue 108
     ##note         the authors translated the codon GAT for residue 337 as
                      Gln
GENETICS
   #gene         CYR61
   #introns      21/3; 93/1; 208/1; 279/3
CLASSIFICATION   #superfamily von Willebrand factor type C repeat homology
FEATURE
   99-166            #domain von Willebrand factor type C repeat homology
                      #label VWC
SUMMARY          #length 379  #molecular-weight 41709  #checksum 3726

  Query Match          14.9%;  Score 121;  DB 2;  Length 379;
  Best Local Similarity  33.3%;  Pred. No. 2.01e-07;
  Matches   19;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db    298 YAGCSSVKKYRPKYC-GSCVDGRCCTPLQTRTVKMRFRCEDGEMFSKNVMMIQSCKC 353
          ||:|:: |   : | |:|   :|| | ::: |  |:| ||  |  :|     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    6
ENTRY            IJFFTM     #type complete
TITLE            cadherin-related tumor suppressor precursor - fruit fly
                  (Drosophila melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                  16-Feb-1997
ACCESSIONS       A41087; B41087
REFERENCE        A41087
   #authors      Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.;
                  Bryant, P.J.; Goodman, C.S.
   #journal      Cell (1991) 67:853-868
   #title        The fat tumor suppressor gene in Drosophila encodes a novel
                  member of the cadherin gene superfamily.
   #cross-references MUID:92069752
   #accession    A41087
```

```
     ##molecule_type mRNA
     ##residues      143-485;1279-5147 ##label MAH
     ##cross-references GB:M80537
   #accession    B41087
     ##molecule_type DNA
     ##residues      1-142;487-1278 ##label MA2
     ##cross-references GB:M80537
     ##note         1229-Gly and 1233-Ser were also found
GENETICS
   #gene         fat
     ##cross-references FlyBase:FBgn0001075
CLASSIFICATION   #superfamily cadherin-related tumor suppressor; cadherin
                  repeat homology; EGF homology
KEYWORDS         calcium binding; cell adhesion; duplication; transmembrane
                  protein
FEATURE
   1-35             #domain signal sequence #status predicted #label SIG\
   36-5147          #product cadherin-related tumor suppressor #status
                     predicted #label MAT\
   36-4583          #domain extracellular #status predicted #label EXT\
   51-156           #domain cadherin repeat homology #label CR1\
   159-270          #domain cadherin repeat homology #label CR2\
   271-382          #domain cadherin repeat homology #label CR3\
   390-494          #domain cadherin repeat homology #label CR4\
   497-599          #domain cadherin repeat homology #label CR5\
   602-708          #domain cadherin repeat homology #label CR6\
   718-822          #domain cadherin repeat homology #label CR7\
   831-942          #domain cadherin repeat homology #label CR8\
   948-1049         #domain cadherin repeat homology #label CR9\
   1052-1153        #domain cadherin repeat homology #label C10\
   1156-1278        #domain cadherin repeat homology #label C11\
   1281-1384        #domain cadherin repeat homology #label C12\
   1387-1489        #domain cadherin repeat homology #label C13\
   1492-1601        #domain cadherin repeat homology #label C14\
   1607-1713        #domain cadherin repeat homology #label C15\
   1717-1823        #domain cadherin repeat homology #label C16\
   1826-1922        #domain cadherin repeat homology #label C17\
   1925-2027        #domain cadherin repeat homology #label C18\
   2028-2167        #domain cadherin repeat homology #label C19\
   2169-2278        #domain cadherin repeat homology #label C20\
   2281-2384        #domain cadherin repeat homology #label C99\
   2387-2491        #domain cadherin repeat homology #label C21\
   2494-2596        #domain cadherin repeat homology #label C22\
   2599-2703        #domain cadherin repeat homology #label C23\
   2707-2810        #domain cadherin repeat homology #label C24\
   2813-2913        #domain cadherin repeat homology #label C25\
   2915-3013        #domain cadherin repeat homology #label C26\
   3014-3124        #domain cadherin repeat homology #label C27\
   3127-3229        #domain cadherin repeat homology #label C28\
   3232-3334        #domain cadherin repeat homology #label C29\
   3337-3439        #domain cadherin repeat homology #label C30\
   3442-3545        #domain cadherin repeat homology #label C31\
   3548-3651        #domain cadherin repeat homology #label C32\
   3654-3756        #domain cadherin repeat homology #label C33\
   3954-4010        #domain EGF homology #label EG1\
   4017-4048        #domain EGF homology #label EG2\
   4056-4089        #domain EGF homology #label EG3\
   4096-4127        #domain EGF homology #label EG4\
   4584-4609        #domain transmembrane #status predicted #label TMM\
   4610-5147        #domain intracellular #status predicted #label INT
SUMMARY          #length 5147  #molecular-weight 564895  #checksum 6994

  Query Match          14.3%;  Score 116;  DB 1;  Length 5147;
  Best Local Similarity  43.3%;  Pred. No. 1.63e-06;
  Matches    13;  Conservative    9;  Mismatches   6;  Indels   2;  Gaps   2;

Db   4067 CQRSPDGSSYFCLCRPGFRGNQCESVSDSC 4096
          |: | :| :| |||:||| |::|:   ::|
Qy      2 CHISDQGEPY-CLCQPGFSGEHCQQ-ENPC 29


RESULT    7
ENTRY           B26637     #type fragment
TITLE           neurogenic repetitive locus 95F protein - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
                  14-Aug-1998
ACCESSIONS      B26637
REFERENCE       A91081
   #authors     Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigel,
                  D.; Vaessin, H.; Campos-Ortega, J.A.
   #journal     EMBO J. (1987) 6:761-766
   #title       EGF homologous sequences encoded in the genome of Drosophila
                  melanogaster, and their relation to neurogenic genes.
   #cross-references MUID:87218537
   #accession   B26637
     ##molecule_type mRNA
     ##residues      1-293 ##label KNU
     ##cross-references GB:X05144; NID:g7519; PID:g929536
GENETICS
   #gene         FlyBase:crb
     ##cross-references FlyBase:FBgn0000368
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         transmembrane protein
FEATURE
   216-252          #domain EGF homology #label EGF
SUMMARY          #length 293  #checksum 3413

  Query Match          13.9%;  Score 113;  DB 2;  Length 293;
  Best Local Similarity  43.8%;  Pred. No. 5.62e-06;
  Matches    14;  Conservative    8;  Mismatches   8;  Indels   2;  Gaps   2;

Db    150 C-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 180
          | |:: :  :| ||||| |:||:|: : |  |
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32


RESULT    8
ENTRY           A35672     #type complete
TITLE           crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A35672
REFERENCE       A35672
   #authors     Tepass, U.; Theres, C.; Knust, E.
   #journal     Cell (1990) 61:787-799
   #title       crumbs encodes an EGF-like protein expressed on apical
                  membranes of Drosophila epithelial cells and required for
                  organization of epithelia.
   #cross-references MUID:90263104
   #accession   A35672
     ##status        preliminary
     ##molecule_type mRNA
     ##residues      1-2139 ##label TEP
     ##cross-references GB:M33753
     ##note         the authors translated the codon GGC for residue 1928 as
                     Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene         FlyBase:crb
     ##cross-references FlyBase:FBgn0000368
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         transmembrane protein
FEATURE
   691-722          #domain EGF homology #label EGF
SUMMARY          #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match          13.9%;  Score 113;  DB 2;  Length 2139;
  Best Local Similarity  43.8%;  Pred. No. 5.62e-06;
  Matches    14;  Conservative    8;  Mismatches   8;  Indels   2;  Gaps   2;

Db   1812 C-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 1842
          | |:: :  :| ||||| |:||:|: : |  |
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32
```

```
RESULT    9
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
   #authors     Bierkamp, C.; Campos-Ortega, J.A.
   #journal     Mech. Dev. (1993) 43:87-100
   #title       A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
   #cross-references MUID:94128602
   #accession   S42612
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2437 ##label BIE
      ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1915-1947         #domain ankyrin repeat homology #label AN1\
   1948-1980         #domain ankyrin repeat homology #label AN2\
   1982-2014         #domain ankyrin repeat homology #label AN3\
   2015-2047         #domain ankyrin repeat homology #label AN4\
   2048-2080         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           13.9%;  Score 113;  DB 2;  Length 2437;
  Best Local Similarity  51.6%;  Pred. No. 5.62e-06;
  Matches   16;  Conservative    5;  Mismatches   7;  Indels   3;  Gaps   3;

Db   1361 C-VSGHLSPRCLCAPGFSGHECQTRMDSPCL 1390
         | :| :  | ||| |||||  || : ::|||
Qy      2 CHISDQGEPYCLCQPGFSGEHCQ-Q-ENPCL 30


RESULT   10
ENTRY           S45306     #type complete
TITLE           notch 3 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S45306
REFERENCE       S45306
   #authors     Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal     Mech. Dev. (1994) 46:123-136
   #title       The novel Notch homologue mouse Notch 3 lacks specific
                  epidermal growth factor-repeats and is expressed in
                  proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession   S45306
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1839-1871         #domain ankyrin repeat homology #label AN1\
   1872-1904         #domain ankyrin repeat homology #label AN2\
   1906-1938         #domain ankyrin repeat homology #label AN3\
   1939-1971         #domain ankyrin repeat homology #label AN4\
   1972-2004         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           13.8%;  Score 112;  DB 2;  Length 2318;
  Best Local Similarity  50.0%;  Pred. No. 8.46e-06;
  Matches   11;  Conservative    8;  Mismatches   2;  Indels   1;  Gaps   1;

Db    949 CLCRPGYTGTHCQYEADPCFSR 970
          |||:||::| ||| : :||:::
Qy     12 CLCQPGFSGEHCQ-QENPCLGQ 32


RESULT   11
ENTRY           S18188     #type complete
TITLE           notch protein homolog - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
                  12-Feb-1999
ACCESSIONS      S18188
REFERENCE       S18188
   #authors     Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal     Development (1991) 113:199-205
   #title       A homolog of Drosophila Notch expressed during mammalian
                  development.
   #cross-references MUID:92111383
   #accession   S18188
      ##molecule_type mRNA
      ##residues     1-2531 ##label WEI
      ##cross-references EMBL:X57405; NID:g57634; PID:g57635
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1917-1949         #domain ankyrin repeat homology #label AN1\
   1950-1982         #domain ankyrin repeat homology #label AN2\
   1984-2016         #domain ankyrin repeat homology #label AN3\
   2017-2049         #domain ankyrin repeat homology #label AN4\
   2050-2082         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2531  #molecular-weight 270907  #checksum 2705

  Query Match           13.7%;  Score 111;  DB 2;  Length 2531;
  Best Local Similarity  35.5%;  Pred. No. 1.27e-05;
  Matches   11;  Conservative   12;  Mismatches   7;  Indels   1;  Gaps   1;

Db     36 RCEVANGTEA-CVCSGAFVGQRCQDPSPCLS 65
            :| :::  |: |:|  :| |::||: :|||:
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLG 31


RESULT   12
ENTRY           A46019     #type complete
TITLE           gene Notch-1 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A46019
REFERENCE       A46019
   #authors     del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins,
                  N.A.; Copeland, N.G.; Gridley, T.
   #journal     Genomics (1993) 15:259-264
   #title       Cloning, analysis, and chromosomal localization of Notch-1, a
                  mouse homolog of Drosophila Notch.
   #cross-references MUID:93194170
   #accession   A46019
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues     1-2531 ##label DEL
      ##cross-references GB:Z11886; GB:S47228; NID:g288502; PID:g288503
      ##note        sequence extracted from NCBI backbone (NCBIP:127318)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   757-788           #domain EGF homology #label EGF\
   1917-1948         #domain ankyrin repeat homology #label AN1\
   1949-1981         #domain ankyrin repeat homology #label AN2\
   1983-2015         #domain ankyrin repeat homology #label AN3\
   2016-2048         #domain ankyrin repeat homology #label AN4\
   2049-2081         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2531  #molecular-weight 271312  #checksum 6611

  Query Match           13.7%;  Score 111;  DB 2;  Length 2531;
  Best Local Similarity  24.2%;  Pred. No. 1.27e-05;
```

```
  Matches   24;  Conservative   28;  Mismatches  39;  Indels   8;  Gaps   8;

Db    589 CLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNSYLCLCLKGTTGPNCEINLDDCASN 648
          ||||||::| ||: : | | :|  |     : |:::|   |  ::  |  |     |::
Qy     12 CLCQPGFSGEHCQQE-NPCLGQVVREV-I-R-RQKGYAS-CATASKVPIMECR-GGCGPQ 65

Db    649 PCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            |::     :   ::|:  :::   : : :: ||:  :|
Qy     66 CCQPTRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCLAC 102


RESULT   13
ENTRY            H71707     #type complete
TITLE            hypothetical protein RP006 - Rickettsia prowazekii
ORGANISM         #formal_name Rickettsia prowazekii
DATE             21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
                   21-Nov-1998
ACCESSIONS       H71707
REFERENCE        A71630
   #authors      Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
                   Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
                   Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
                   C.G.
   #journal      Nature (1998) 396:133-140
   #title        The genome sequence of Rickettsia prowazekii and the origin
                   of mitochondria.
   #accession    H71707
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-706 ##label AND
      ##cross-references GB:AJ235270; GB:AJ235269; NID:g3860572; PID:e1342322;
                          PID:g3860578
      ##experimental_source strain Madrid E
GENETICS
   #gene         RP006
SUMMARY          #length 706  #molecular-weight 80294  #checksum 4323

  Query Match           13.5%;  Score 110;  DB 2;  Length 706;
  Best Local Similarity  36.6%;  Pred. No. 1.91e-05;
  Matches   15;  Conservative   13;  Mismatches   9;  Indels   4;  Gaps   3;

Db    649 SLMQCMMKGICG-QCIQKVKGKQ-KYIFACSEQNQNVEIID 687
          ::|:|  :| || || |  ::|: ||:| |:: :  || ::
Qy     54 PIMEC--RGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVE 92


RESULT   14
ENTRY            A49128     #type complete
TITLE            cell-fate determining gene Notch2 protein - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                   14-Aug-1998
ACCESSIONS       A49128
REFERENCE        A49128
   #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal      Development (1992) 116:931-941
   #title        Notch2: a second mammalian Notch gene.
   #cross-references MUID:93202015
   #accession    A49128
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues      1-2471 ##label WEI
      ##experimental_source Schwann cell
      ##note          sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   1029-1060          #domain EGF homology #label EGF\
   1876-1908          #domain ankyrin repeat homology #label AN1\
   1909-1941          #domain ankyrin repeat homology #label AN2\
   1943-1975          #domain ankyrin repeat homology #label AN3\
   1976-2008          #domain ankyrin repeat homology #label AN4\
   2009-2041          #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           13.5%;  Score 110;  DB 2;  Length 2471;
  Best Local Similarity  46.4%;  Pred. No. 1.91e-05;
  Matches   13;  Conservative    2;  Mismatches  13;  Indels   0;  Gaps   0;

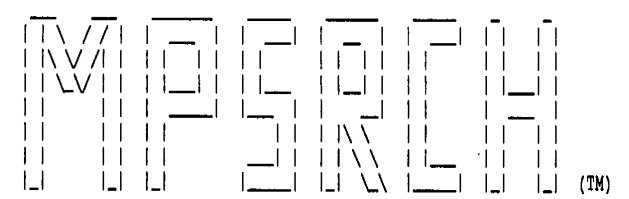
Db     41 CVTYHNGTGYCRCPEGFLGEYCQHRDPC 68
            |     |  || |  || || ||: :||
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQENPC 29


RESULT   15
ENTRY            A49175     #type fragment
TITLE            Motch B protein - mouse (fragment)
ALTERNATE_NAMES  Notch homolog
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                   14-Aug-1998
ACCESSIONS       A49175; PH1570; S32113
REFERENCE        A49175
   #authors      Lardelli, M.; Lendahl, U.
   #journal      Exp. Cell Res. (1993) 204:364-372
   #title        Motch A and Motch B--two mouse Notch homologues coexpressed
                   in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession    A49175
      ##status        preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-1203 ##label LAR
      ##cross-references EMBL:X68279; NID:g287989; PID:g287990
      ##experimental_source embryo
      ##note          sequence extracted from NCBI backbone (NCBIP:126158)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermaland neural fate for cells in the early
             embryo.
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   560-591            #domain EGF homology #label EGF
SUMMARY          #length 1203  #checksum 910

  Query Match           13.3%;  Score 108;  DB 2;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 4.29e-05;
  Matches   16;  Conservative    6;  Mismatches   6;  Indels   4;  Gaps   4;

Db    571 CH-NTQG-SYVCECPPGFSGMDCEEDINDCLA 600
          || : || :| | | |||||  |::: | ||:
Qy      2 CHISDQGEPY-CLCQPGFSGEHCQQE-NPCLG 31


Search completed: Fri May 28 08:59:30 1999
Job time : 21 secs.
```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 08:59:48 1999; MasPar time 5.03 Seconds
578.731 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-6
Description: (1-103) from US09191647.pep
Perfect Score: 813
Sequence: 1 QCHISDQGEPYCLCQPGFSG.....GSSFVEEVERHLECGCLACS 103

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 37.938; Variance 55.390; scale 0.685

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	139	17.1	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	8.76e-13
2	129	15.9	375	1	CE10_CHICK	CEF-10 PROTEIN PRECURS	1.31e-10
3	121	14.9	379	1	CYR6_MOUSE	CYR61 PROTEIN PRECURSO	6.52e-09
4	116	14.3	381	1	CYR6_HUMAN	CYR61 PROTEIN PRECURSO	7.11e-08
5	116	14.3	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	7.11e-08
6	113	13.9	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	2.92e-07
7	113	13.9	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	2.92e-07
8	112	13.8	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	4.65e-07
9	111	13.7	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	7.40e-07
10	111	13.7	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	7.40e-07
11	110	13.5	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	1.18e-06
12	108	13.3	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	2.95e-06
13	105	12.9	2813	1	VWF_HUMAN	VON WILLEBRAND FACTOR	1.16e-05
14	105	12.9	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	1.16e-05
15	105	12.9	4543	1	LRP1_CHICK	LOW-DENSITY LIPOPROTEI	1.16e-05
16	103	12.7	2415	1	PGCA_HUMAN	AGGRECAN CORE PROTEIN	2.84e-05
17	103	12.7	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	2.84e-05
18	102	12.5	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	4.44e-05
19	102	12.5	603	1	FA12_CAVPO	COAGULATION FACTOR XII	4.44e-05
20	102	12.5	615	1	FA12_HUMAN	COAGULATION FACTOR XII	4.44e-05
21	102	12.5	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	4.44e-05
22	101	12.4	1056	1	MUC5_HUMAN	TRACHEOBRONCHIAL MUCIN	6.92e-05
23	101	12.4	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	6.92e-05
24	100	12.3	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	1.08e-04
25	99	12.2	427	1	MFGM_RAT	MILK FAT GLOBULE-EGF F	1.67e-04
26	96	11.8	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	6.14e-04
27	96	11.8	463	1	MFGM_MOUSE	MILK FAT GLOBULE-EGF F	6.14e-04
28	96	11.8	3672	1	LML2_CAEEL	LAMININ-LIKE PROTEIN K	6.14e-04
29	96	11.8	3712	1	LMA_DROME	LAMININ ALPHA CHAIN PR	6.14e-04
30	95	11.7	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	9.44e-04
31	95	11.7	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	9.44e-04
32	95	11.7	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	9.44e-04
33	94	11.6	84	1	HBGF_PIG	HEPARIN-BINDING EGF-LI	1.45e-03
34	94	11.6	208	1	HBGF_HUMAN	HEPARIN-BINDING EGF-LI	1.45e-03
35	94	11.6	208	1	HBGF_CERAE	HEPARIN-BINDING EGF-LI	1.45e-03
36	92	11.3	611	1	LEM2_CANFA	E-SELECTIN PRECURSOR (	3.37e-03
37	92	11.3	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	3.37e-03
38	92	11.3	4544	1	LRP1_HUMAN	LOW-DENSITY LIPOPROTEI	3.37e-03
39	91	11.2	121	1	TGFA_MACMU	TRANSFORMING GROWTH FA	5.12e-03
40	91	11.2	160	1	TGFA_PIG	TRANSFORMING GROWTH FA	5.12e-03
41	91	11.2	160	1	TGFA_HUMAN	TRANSFORMING GROWTH FA	5.12e-03
42	91	11.2	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	5.12e-03
43	91	11.2	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	5.12e-03
44	91	11.2	1168	1	LMB3_MOUSE	LAMININ BETA-3 CHAIN P	5.12e-03
45	91	11.2	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	5.12e-03

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1     36
FT   CHAIN       37   1480       SLIT PROTEIN.
FT   DOMAIN      70    104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105    230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231    294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295    326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327    452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453    518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519    550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551    653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654    714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715    746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747    848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849    910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105    115       LRR 1-1.
FT   REPEAT     116    139       LRR 1-2.
FT   REPEAT     140    163       LRR 1-3.
FT   REPEAT     164    187       LRR 1-4.
FT   REPEAT     188    211       LRR 1-5.
FT   REPEAT     212    230       LRR 1-6.
FT   REPEAT     327    337       LRR 2-1.
FT   REPEAT     338    361       LRR 2-2.
FT   REPEAT     362    385       LRR 2-3.
FT   REPEAT     386    409       LRR 2-4.
FT   REPEAT     410    433       LRR 2-5.
FT   REPEAT     434    452       LRR 2-6.
FT   REPEAT     551    562       LRR 3-1.
FT   REPEAT     563    586       LRR 3-2.
FT   REPEAT     587    610       LRR 3-3.
FT   REPEAT     611    634       LRR 3-4.
FT   REPEAT     635    653       LRR 3-5.
FT   REPEAT     747    757       LRR 4-1.
FT   REPEAT     758    781       LRR 4-2.
FT   REPEAT     782    805       LRR 4-3.
FT   REPEAT     806    829       LRR 4-4.
FT   REPEAT     830    848       LRR 4-5.
FT   DOMAIN     907    944       EGF-LIKE 1.
FT   DOMAIN     946    983       EGF-LIKE 2.
FT   DOMAIN     985   1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024   1062       EGF-LIKE 4.
FT   DOMAIN    1064   1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111   1149       EGF-LIKE 6.
FT   DOMAIN    1353   1392       EGF-LIKE 7.
FT   DOMAIN    1409   1480       CTCK.
FT   VARSPLIC  1394   1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111    111       POTENTIAL.
FT   CARBOHYD   207    207       POTENTIAL.
FT   CARBOHYD   357    357       POTENTIAL.
FT   CARBOHYD   435    435       POTENTIAL.
FT   CARBOHYD   783    783       POTENTIAL.
FT   CARBOHYD   788    788       POTENTIAL.
FT   CARBOHYD   958    958       POTENTIAL.
FT   CARBOHYD   998    998       POTENTIAL.
FT   CARBOHYD  1060   1060       POTENTIAL.
FT   CARBOHYD  1159   1159       POTENTIAL.
FT   CARBOHYD  1175   1175       POTENTIAL.
FT   CARBOHYD  1243   1243       POTENTIAL.
FT   CARBOHYD  1292   1292       POTENTIAL.
FT   DISULFID   911    922       BY SIMILARITY.
FT   DISULFID   916    932       BY SIMILARITY.
FT   DISULFID   934    943       BY SIMILARITY.
FT   DISULFID   950    961       BY SIMILARITY.
FT   DISULFID   955    971       BY SIMILARITY.
FT   DISULFID   973    982       BY SIMILARITY.
FT   DISULFID   989   1001       BY SIMILARITY.
FT   DISULFID   995   1010       BY SIMILARITY.
FT   DISULFID  1012   1021       BY SIMILARITY.
FT   DISULFID  1028   1041       BY SIMILARITY.
FT   DISULFID  1035   1050       BY SIMILARITY.
FT   DISULFID  1052   1061       BY SIMILARITY.
FT   DISULFID  1068   1079       BY SIMILARITY.
FT   DISULFID  1073   1088       BY SIMILARITY.
FT   DISULFID  1090   1099       BY SIMILARITY.
FT   DISULFID  1115   1125       BY SIMILARITY.
FT   DISULFID  1120   1137       BY SIMILARITY.
FT   DISULFID  1139   1148       BY SIMILARITY.
FT   DISULFID  1357   1368       BY SIMILARITY.
FT   DISULFID  1362   1380       BY SIMILARITY.
FT   DISULFID  1382   1391       BY SIMILARITY.
FT   DISULFID  1409   1443       BY SIMILARITY.
FT   DISULFID  1423   1457       BY SIMILARITY.
FT   DISULFID  1434   1473       BY SIMILARITY.
FT   DISULFID  1438   1475       BY SIMILARITY.
FT   DISULFID  1442   1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          17.1%;  Score 139;  DB 1;  Length 1480;
  Best Local Similarity  35.7%;  Pred. No. 8.76e-13;
  Matches   15;  Conservative    9;  Mismatches  18;  Indels   0;  Gaps   0;

Db   1434 CVGGCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGC 1475
          | |||| ||| :  :|||   |::  :: :::   |||
Qy     58 CRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT   2
ID   CE10_CHICK     STANDARD;      PRT;   375 AA.
AC   P19336;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CEF-10 PROTEIN PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89145206.
RA   SIMMONS D.L., LEVY D.B., YANNONI Y., ERIKSON R.L.;
RT   "Identification of a phorbol ester-repressible v-src-inducible gene.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:1178-1182(1989).
CC   -!- FUNCTION: PROBABLE SECRETED REGULATORY PROTEIN.
CC   -!- INDUCTION: BY V-SRC.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04496; G211436; -.
DR   PIR; A41428; A41428.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
```

```
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1     22
FT   CHAIN       23    375       CEF-10 PROTEIN.
FT   DOMAIN      98    164       VWFC.
FT   DOMAIN     281    355       CTCK.
FT   DISULFID   281    318       BY SIMILARITY.
FT   DISULFID   298    332       BY SIMILARITY.
FT   DISULFID   309    348       BY SIMILARITY.
FT   DISULFID   312    350       BY SIMILARITY.
FT   DISULFID   317    354       BY SIMILARITY.
SQ   SEQUENCE   375 AA;  40651 MW;  68B4BC92 CRC32;

  Query Match           15.9%;  Score 129;  DB 1;  Length 375;
  Best Local Similarity  33.3%;  Pred. No. 1.31e-10;
  Matches   19;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db    295 YAGCSSVKKYRPKYC-GSCVDGRCCTPQQTRTVKIRFRCDDGETFTKSVMMIQSCRC 350
          ||:|::  |   : | |:|   :|| | :::  |  |:| || :|   |     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    3
ID   CYR6_MOUSE     STANDARD;      PRT;   379 AA.
AC   P18406;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (3CH61).
GN   IGFBP10 OR CYR61.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=FIBROBLAST;
RX   MEDLINE; 90287146.
RA   O'BRIEN T.P., YANG G.P., SANDERS L., LAU L.F.;
RT   "Expression of cyr61, a growth factor-inducible immediate-early
RT   gene.";
RL   MOL. CELL. BIOL. 10:3569-3577(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AJ; TISSUE=EMBRYONIC FIBROBLAST;
RX   MEDLINE; 91288203.
RA   LATINKIC B.V., O'BRIEN T.P., LAU L.F.;
RT   "Promoter function and structure of the growth factor-inducible
RT   immediate early gene cyr61.";
RL   NUCLEIC ACIDS RES. 19:3261-3267(1991).
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION.
CC   -!- TISSUE SPECIFICITY: LOW IN KIDNEY, ADRENAL GLAND, TESTES, BRAIN,
CC       AND OVARY, MODERATE IN HEART, UTERUS, AND SKELETAL MUSCLE, HIGHEST
CC       IN LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED FROM G(0)/G(1) THROUGH MID-G(1) IN
CC       NORMAL CELLS, AND AT A CONSTANT LEVEL IN RAPIDLY GROWING CELLS.
CC   -!- INDUCTION: BY GROWTH FACTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M32490; G309206; -.
DR   EMBL; X56790; G50633; -.
DR   PIR; A35669; A35669.
DR   MGD; MGI:88613; IGFBP10.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    379       CYR61 PROTEIN.
FT   DOMAIN      98    164       VWFC.
FT   DOMAIN     284    358       CTCK.
FT   DISULFID   284    321       BY SIMILARITY.
FT   DISULFID   301    335       BY SIMILARITY.
FT   DISULFID   312    351       BY SIMILARITY.
FT   DISULFID   315    353       BY SIMILARITY.
FT   DISULFID   320    357       BY SIMILARITY.
SQ   SEQUENCE   379 AA;  41709 MW;  116B80C7 CRC32;

  Query Match           14.9%;  Score 121;  DB 1;  Length 379;
  Best Local Similarity  33.3%;  Pred. No. 6.52e-09;
  Matches   19;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db    298 YAGCSSVKKYRPKYC-GSCVDGRCCTPLQTRTVKMRFRCEDGEMFSKNVMMIQSCKC 353
          ||:|::  |   : | |:|   :|| | :::  |  |:| ||  |  :|     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    4
ID   CYR6_HUMAN     STANDARD;      PRT;   381 AA.
AC   O00622; O14934;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (GIG1 PROTEIN) (INSULIN-LIKE GROWTH FACTOR-
DE   BINDING PROTEIN 10).
GN   IGFBP10 OR CYR61 OR GIG1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALBRECHT C., VON DER KAMMER H., KLAUDINY J., MAYHAUS M., NITSCH R.M.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97280750.
RA   JAY P., BERGE-LEFRANC J.L., MARSOLLIER C., MEJEAN C., TAVIAUX S.,
RA   BERTA P.;
RT   "The human growth factor-inducible immediate early gene, CYR61, maps
RT   to chromosome 1p.";
RL   ONCOGENE 14:1753-1757(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   KOLESNIKOVA T.V., LAU L.F.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   BI A.B., YU L.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12084; E311857; -.
DR   EMBL; U62015; G2130527; -.
DR   EMBL; AF003594; G2196782; -.
DR   EMBL; AF031385; G2606094; -.
DR   MIM; 602369; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25   381       CYR61 PROTEIN.
FT   DOMAIN      98   164       VWFC.
FT   DOMAIN     286   360       CTCK.
FT   DISULFID   286   323       BY SIMILARITY.
FT   DISULFID   303   337       BY SIMILARITY.
FT   DISULFID   314   353       BY SIMILARITY.
FT   DISULFID   317   355       BY SIMILARITY.
FT   DISULFID   322   359       BY SIMILARITY.
FT   CONFLICT   210   210       L -> I (IN REF. 4).
FT   CONFLICT   220   220       L -> R (IN REF. 4).
SQ   SEQUENCE   381 AA;  42026 MW;  2B091D9E CRC32;

  Query Match           14.3%;  Score 116;  DB 1;  Length 381;
  Best Local Similarity  33.3%;  Pred. No. 7.11e-08;
  Matches   19;  Conservative   10;  Mismatches  26;  Indels   2;  Gaps   2;

Db    300 YAGCLSVKKYRPKYC-GSCVDGRCCTPQLTRTVKMRFRCEDGETFSKNVMMIQSCKC 355
          ||:| :  |   : | |:|   :|| |  ::  |  |:| || :|  :|     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT   5
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN   FT.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92069752.
RA   MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA   GOODMAN C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   CELL 67:853-868(1991).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC       TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC       DIFFERENTIATION AND MORPHOGENESIS, AND DEAT DURING THE PUPAL
CC       STAGE.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY; CONTAINS 37 CADHERINS-TYPE REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80537; G157409; -.
DR   PIR; A41087; IJFFTM.
DR   FLYBASE; FBgn0001075; ft.
DR   PROSITE; PS00232; CADHERIN; 22.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PFAM; PF00008; EGF; 4.
DR   PFAM; PF00028; cadherin; 33.
DR   PFAM; PF00054; laminin_G; 1.
DR   HSSP; P00740; 1IXA.
KW   CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW   CALCIUM-BINDING; REPEAT; EGF-LIKE DOMAIN.
FT   SIGNAL       1    35       POTENTIAL.
FT   CHAIN       36  5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN      36  4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  4584  4609       POTENTIAL.
FT   DOMAIN    4610  5147       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      36   156       CADHERIN 1.
FT   REPEAT     157   270       CADHERIN 2.
FT   REPEAT     271   382       CADHERIN 3.
FT   REPEAT     383   494       CADHERIN 4.
FT   REPEAT     495   599       CADHERIN 5.
FT   REPEAT     600   708       CADHERIN 6.
FT   REPEAT     709   820       CADHERIN 7.
FT   REPEAT     821   942       CADHERIN 8.
FT   REPEAT     943  1049       CADHERIN 9.
FT   REPEAT    1050  1153       CADHERIN 10.
FT   REPEAT    1154  1278       CADHERIN 11.
FT   REPEAT    1279  1384       CADHERIN 12.
FT   REPEAT    1385  1489       CADHERIN 13.
FT   REPEAT    1490  1601       CADHERIN 14.
FT   REPEAT    1602  1713       CADHERIN 15.
FT   REPEAT    1714  1823       CADHERIN 16.
FT   REPEAT    1824  1922       CADHERIN 17.
FT   REPEAT    1923  2027       CADHERIN 18.
FT   REPEAT    2028  2167       CADHERIN 19.
FT   REPEAT    2168  2278       CADHERIN 20.
FT   REPEAT    2279  2385       CADHERIN 21.
FT   REPEAT    2386  2491       CADHERIN 22.
FT   REPEAT    2492  2596       CADHERIN 23.
FT   REPEAT    2597  2703       CADHERIN 24.
FT   REPEAT    2704  2810       CADHERIN 25.
FT   REPEAT    2811  2913       CADHERIN 26.
FT   REPEAT    2914  3013       CADHERIN 27.
FT   REPEAT    3014  3124       CADHERIN 28.
FT   REPEAT    3125  3229       CADHERIN 29.
FT   REPEAT    3230  3334       CADHERIN 30.
FT   REPEAT    3335  3439       CADHERIN 31.
FT   REPEAT    3440  3545       CADHERIN 32.
FT   REPEAT    3546  3651       CADHERIN 33.
FT   REPEAT    3652  3756       CADHERIN 34.
FT   DOMAIN    3950  4011       EGF-LIKE 1.
FT   DOMAIN    4013  4049       EGF-LIKE 2.
FT   DOMAIN    4052  4090       EGF-LIKE 3.
FT   DOMAIN    4092  4128       EGF-LIKE 4.
FT   DOMAIN    4321  4362       EGF-LIKE 5.
FT   DISULFID  3954  3966       BY SIMILARITY.
FT   DISULFID  3960  3999       BY SIMILARITY.
FT   DISULFID  4001  4010       BY SIMILARITY.
FT   DISULFID  4017  4028       BY SIMILARITY.
FT   DISULFID  4022  4037       BY SIMILARITY.
FT   DISULFID  4039  4048       BY SIMILARITY.
FT   DISULFID  4056  4067       BY SIMILARITY.
FT   DISULFID  4061  4078       BY SIMILARITY.
FT   DISULFID  4080  4089       BY SIMILARITY.
FT   DISULFID  4096  4107       BY SIMILARITY.
```

```
FT   DISULFID   4101   4116       BY SIMILARITY.
FT   DISULFID   4118   4127       BY SIMILARITY.
FT   DISULFID   4325   4341       BY SIMILARITY.
FT   DISULFID   4334   4350       BY SIMILARITY.
FT   DISULFID   4352   4361       BY SIMILARITY.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    257    257       POTENTIAL.
FT   CARBOHYD    276    276       POTENTIAL.
FT   CARBOHYD    280    280       POTENTIAL.
FT   CARBOHYD    402    402       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    605    605       POTENTIAL.
FT   CARBOHYD    631    631       POTENTIAL.
FT   CARBOHYD   1155   1155       POTENTIAL.
FT   CARBOHYD   1367   1367       POTENTIAL.
FT   CARBOHYD   1458   1458       POTENTIAL.
FT   CARBOHYD   1751   1751       POTENTIAL.
FT   CARBOHYD   1831   1831       POTENTIAL.
FT   CARBOHYD   1880   1880       POTENTIAL.
FT   CARBOHYD   2080   2080       POTENTIAL.
FT   CARBOHYD   2171   2171       POTENTIAL.
FT   CARBOHYD   2247   2247       POTENTIAL.
FT   CARBOHYD   2290   2290       POTENTIAL.
FT   CARBOHYD   2437   2437       POTENTIAL.
FT   CARBOHYD   2581   2581       POTENTIAL.
FT   CARBOHYD   2799   2799       POTENTIAL.
FT   CARBOHYD   2920   2920       POTENTIAL.
FT   CARBOHYD   2946   2946       POTENTIAL.
FT   CARBOHYD   2967   2967       POTENTIAL.
FT   CARBOHYD   3167   3167       POTENTIAL.
FT   CARBOHYD   3303   3303       POTENTIAL.
FT   CARBOHYD   3386   3386       POTENTIAL.
FT   CARBOHYD   3389   3389       POTENTIAL.
FT   CARBOHYD   3525   3525       POTENTIAL.
FT   CARBOHYD   3852   3852       POTENTIAL.
FT   CARBOHYD   3865   3865       POTENTIAL.
FT   CARBOHYD   3905   3905       POTENTIAL.
FT   CARBOHYD   4306   4306       POTENTIAL.
FT   CARBOHYD   4414   4414       POTENTIAL.
FT   CARBOHYD   4471   4471       POTENTIAL.
FT   CARBOHYD   4487   4487       POTENTIAL.
FT   CARBOHYD   4539   4539       POTENTIAL.
FT   CARBOHYD   4550   4550       POTENTIAL.
FT   VARIANT    1229   1229       S -> G.
FT   VARIANT    1233   1233       G -> S.
     SEQUENCE   5147 AA;  564868 MW;  1EF20E13 CRC32;

  Query Match          14.3%;  Score 116;  DB 1;  Length 5147;
  Best Local Similarity  43.3%;  Pred. No. 7.11e-08;
  Matches   13;  Conservative    9;  Mismatches   6;  Indels   2;  Gaps   2;

Db   4067 CQRSPDGSSYFCLCRPGFRGNQCESVSDSC 4096
          |: | :| :| |||:||| |::|:   ::|
Qy      2 CHISDQGEPY-CLCQPGFSGEHCQQ-ENPC 29


RESULT    6
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL        1     90
FT   CHAIN        91   2139       CRUMBS PROTEIN.
FT   DOMAIN       91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2085   2111       POTENTIAL.
FT   DOMAIN     2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      267    303       EGF-LIKE 1.
FT   DOMAIN      306    343       EGF-LIKE 2.
FT   DOMAIN      348    386       EGF-LIKE 3.
FT   DOMAIN      388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      427    463       EGF-LIKE 5.
FT   DOMAIN      464    500       EGF-LIKE 6.
FT   DOMAIN      501    532       EGF-LIKE 7.
FT   DOMAIN      545    581       EGF-LIKE 8.
FT   DOMAIN      582    611       EGF-LIKE 9.
FT   DOMAIN      609    646       EGF-LIKE 10.
FT   DOMAIN      648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      840    902       EGF-LIKE 16.
FT   DOMAIN      904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      980   1021       EGF-LIKE 19.
FT   DOMAIN     1207   1243       EGF-LIKE 20.
FT   DOMAIN     1481   1517       EGF-LIKE 21.
FT   DOMAIN     1759   1795       EGF-LIKE 22.
FT   DOMAIN     1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1874   1915       EGF-LIKE 25.
FT   DOMAIN     1915   1951       EGF-LIKE 26.
```

```
FT   DOMAIN     1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     2030   2070       EGF-LIKE 29.
FT   DISULFID    271    282       BY SIMILARITY.
FT   DISULFID    276    291       BY SIMILARITY.
FT   DISULFID    293    302       BY SIMILARITY.
FT   DISULFID    310    321       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    333    342       BY SIMILARITY.
FT   DISULFID    352    363       BY SIMILARITY.
FT   DISULFID    357    374       BY SIMILARITY.
FT   DISULFID    376    385       BY SIMILARITY.
FT   DISULFID    392    403       BY SIMILARITY.
FT   DISULFID    397    412       BY SIMILARITY.
FT   DISULFID    414    424       BY SIMILARITY.
FT   DISULFID    431    442       BY SIMILARITY.
FT   DISULFID    436    451       BY SIMILARITY.
FT   DISULFID    453    462       BY SIMILARITY.
FT   DISULFID    468    479       BY SIMILARITY.
FT   DISULFID    473    488       BY SIMILARITY.
FT   DISULFID    490    499       BY SIMILARITY.
FT   DISULFID    505    515       BY SIMILARITY.
FT   DISULFID    509    520       BY SIMILARITY.
FT   DISULFID    522    531       BY SIMILARITY.
FT   DISULFID    549    562       BY SIMILARITY.
FT   DISULFID    556    569       BY SIMILARITY.
FT   DISULFID    571    580       BY SIMILARITY.
FT   DISULFID    586    597       BY SIMILARITY.
FT   DISULFID    591    602       BY SIMILARITY.
FT   DISULFID    604    610       BY SIMILARITY.
FT   DISULFID    613    624       BY SIMILARITY.
FT   DISULFID    618    634       BY SIMILARITY.
FT   DISULFID    636    645       BY SIMILARITY.
FT   DISULFID    652    664       BY SIMILARITY.
FT   DISULFID    659    673       BY SIMILARITY.
FT   DISULFID    675    684       BY SIMILARITY.
FT   DISULFID    691    702       BY SIMILARITY.
FT   DISULFID    696    711       BY SIMILARITY.
FT   DISULFID    713    722       BY SIMILARITY.
FT   DISULFID    729    740       BY SIMILARITY.
FT   DISULFID    734    749       BY SIMILARITY.
FT   DISULFID    751    760       BY SIMILARITY.
FT   DISULFID    767    778       BY SIMILARITY.
FT   DISULFID    772    787       BY SIMILARITY.
FT   DISULFID    789    799       BY SIMILARITY.
FT   DISULFID    806    817       BY SIMILARITY.
FT   DISULFID    811    826       BY SIMILARITY.
FT   DISULFID    828    837       BY SIMILARITY.
FT   DISULFID    844    855       BY SIMILARITY.
FT   DISULFID    849    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID    908    919       BY SIMILARITY.
FT   DISULFID    913    928       BY SIMILARITY.
FT   DISULFID    930    939       BY SIMILARITY.
FT   DISULFID    946    957       BY SIMILARITY.
FT   DISULFID    952    966       BY SIMILARITY.
FT   DISULFID    968    977       BY SIMILARITY.
FT   DISULFID    984    995       BY SIMILARITY.
FT   DISULFID    989   1009       BY SIMILARITY.
FT   DISULFID   1011   1020       BY SIMILARITY.
FT   DISULFID   1211   1222       BY SIMILARITY.
FT   DISULFID   1216   1231       BY SIMILARITY.
FT   DISULFID   1233   1242       BY SIMILARITY.
FT   DISULFID   1485   1496       BY SIMILARITY.
FT   DISULFID   1490   1505       BY SIMILARITY.
FT   DISULFID   1507   1516       BY SIMILARITY.
FT   DISULFID   1763   1774       BY SIMILARITY.
FT   DISULFID   1768   1783       BY SIMILARITY.
FT   DISULFID   1785   1794       BY SIMILARITY.
FT   DISULFID   1801   1812       BY SIMILARITY.
FT   DISULFID   1806   1821       BY SIMILARITY.
FT   DISULFID   1823   1832       BY SIMILARITY.
FT   DISULFID   1839   1850       BY SIMILARITY.
FT   DISULFID   1844   1859       BY SIMILARITY.
FT   DISULFID   1861   1870       BY SIMILARITY.
FT   DISULFID   1878   1889       BY SIMILARITY.
FT   DISULFID   1883   1903       BY SIMILARITY.
FT   DISULFID   1905   1914       BY SIMILARITY.
FT   DISULFID   1919   1930       BY SIMILARITY.
FT   DISULFID   1924   1939       BY SIMILARITY.
FT   DISULFID   1941   1950       BY SIMILARITY.
FT   DISULFID   1957   1968       BY SIMILARITY.
FT   DISULFID   1962   1977       BY SIMILARITY.
FT   DISULFID   1979   1988       BY SIMILARITY.
FT   DISULFID   1995   2008       BY SIMILARITY.
FT   DISULFID   2002   2017       BY SIMILARITY.
FT   DISULFID   2019   2028       BY SIMILARITY.
FT   CARBOHYD     37     37       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    238    238       POTENTIAL.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    336    336       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CARBOHYD    550    550       POTENTIAL.
FT   CARBOHYD    565    565       POTENTIAL.
FT   CARBOHYD    736    736       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    860    860       POTENTIAL.
FT   CARBOHYD    884    884       POTENTIAL.
FT   CARBOHYD    976    976       POTENTIAL.
FT   CARBOHYD   1102   1102       POTENTIAL.
FT   CARBOHYD   1114   1114       POTENTIAL.
FT   CARBOHYD   1138   1138       POTENTIAL.
FT   CARBOHYD   1192   1192       POTENTIAL.
FT   CARBOHYD   1245   1245       POTENTIAL.
FT   CARBOHYD   1255   1255       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   CARBOHYD   1363   1363       POTENTIAL.
FT   CARBOHYD   1441   1441       POTENTIAL.
FT   CARBOHYD   1454   1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          13.9%;  Score 113;  DB 1;  Length 2139;
  Best Local Similarity  43.8%;  Pred. No. 2.92e-07;
  Matches   14;  Conservative    8;  Mismatches   8;  Indels   2;  Gaps   2;

Db   1812 C-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 1842
          | |:: :  :| ||||| |:||:|: : |  |
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32


RESULT    7
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
```

```
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    20       POTENTIAL.
FT   CHAIN       21  2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21  1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725  1747       POTENTIAL.
FT   DOMAIN    1748  2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21    57       EGF-LIKE 1.
FT   DOMAIN      58    98       EGF-LIKE 2.
FT   DOMAIN     101   138       EGF-LIKE 3.
FT   DOMAIN     139   175       EGF-LIKE 4.
FT   DOMAIN     177   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217   254       EGF-LIKE 6.
FT   DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371   409       EGF-LIKE 10.
FT   DOMAIN     411   449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451   487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489   524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526   562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564   599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     601   637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639   674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     676   712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     714   749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     751   787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     789   825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     827   865       EGF-LIKE 22.
FT   DOMAIN     867   903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     905   941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     943   979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     981  1017       EGF-LIKE 26.
FT   DOMAIN    1019  1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1057  1093       EGF-LIKE 28.
FT   DOMAIN    1095  1141       EGF-LIKE 29.
FT   DOMAIN    1143  1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1181  1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1219  1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1265  1303       EGF-LIKE 33.
FT   DOMAIN    1305  1344       EGF-LIKE 34.
FT   DOMAIN    1346  1382       EGF-LIKE 35.
FT   DOMAIN    1385  1423       EGF-LIKE 36.
FT   DOMAIN    1446  1561       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446  1486       LIN/NOTCH 1.
FT   REPEAT    1487  1520       LIN/NOTCH 2.
FT   REPEAT    1521  1561       LIN/NOTCH 3.
FT   DOMAIN    1861  2074       6 X ANK MOTIF REPEATS.
FT   REPEAT    1861  1891       ANK MOTIF 1.
FT   REPEAT    1892  1940       ANK MOTIF 1.
FT   REPEAT    1941  1974       ANK MOTIF 1.
FT   REPEAT    1975  2007       ANK MOTIF 1.
FT   REPEAT    2008  2040       ANK MOTIF 1.
FT   REPEAT    2041  2074       ANK MOTIF 1.
FT   DOMAIN    2265  2276       POLY-GLN (OPA-REPEAT).
FT   DISULFID    25    35       BY SIMILARITY.
FT   DISULFID    29    45       BY SIMILARITY.
FT   DISULFID    47    56       BY SIMILARITY.
FT   DISULFID    62    73       BY SIMILARITY.
FT   DISULFID    67    86       BY SIMILARITY.
FT   DISULFID    88    97       BY SIMILARITY.
FT   DISULFID   105   116       BY SIMILARITY.
FT   DISULFID   110   126       BY SIMILARITY.
FT   DISULFID   128   137       BY SIMILARITY.
FT   DISULFID   143   154       BY SIMILARITY.
FT   DISULFID   148   163       BY SIMILARITY.
FT   DISULFID   165   174       BY SIMILARITY.
FT   DISULFID   181   194       BY SIMILARITY.
FT   DISULFID   188   203       BY SIMILARITY.
FT   DISULFID   205   214       BY SIMILARITY.
FT   DISULFID   221   232       BY SIMILARITY.
FT   DISULFID   226   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   265   280       BY SIMILARITY.
FT   DISULFID   282   291       BY SIMILARITY.
FT   DISULFID   298   311       BY SIMILARITY.
FT   DISULFID   305   320       BY SIMILARITY.
FT   DISULFID   322   331       BY SIMILARITY.
FT   DISULFID   338   349       BY SIMILARITY.
FT   DISULFID   343   358       BY SIMILARITY.
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   375   386       BY SIMILARITY.
FT   DISULFID   380   397       BY SIMILARITY.
FT   DISULFID   399   408       BY SIMILARITY.
FT   DISULFID   415   428       BY SIMILARITY.
FT   DISULFID   422   437       BY SIMILARITY.
FT   DISULFID   439   448       BY SIMILARITY.
FT   DISULFID   455   466       BY SIMILARITY.
FT   DISULFID   460   475       BY SIMILARITY.
FT   DISULFID   477   486       BY SIMILARITY.
FT   DISULFID   493   503       BY SIMILARITY.
FT   DISULFID   498   512       BY SIMILARITY.
FT   DISULFID   514   523       BY SIMILARITY.
FT   DISULFID   530   541       BY SIMILARITY.
FT   DISULFID   535   550       BY SIMILARITY.
FT   DISULFID   552   561       BY SIMILARITY.
FT   DISULFID   568   578       BY SIMILARITY.
FT   DISULFID   573   587       BY SIMILARITY.
FT   DISULFID   589   598       BY SIMILARITY.
FT   DISULFID   605   616       BY SIMILARITY.
FT   DISULFID   610   625       BY SIMILARITY.
FT   DISULFID   627   636       BY SIMILARITY.
FT   DISULFID   643   653       BY SIMILARITY.
FT   DISULFID   648   662       BY SIMILARITY.
FT   DISULFID   664   673       BY SIMILARITY.
FT   DISULFID   680   691       BY SIMILARITY.
FT   DISULFID   685   700       BY SIMILARITY.
FT   DISULFID   702   711       BY SIMILARITY.
FT   DISULFID   718   728       BY SIMILARITY.
FT   DISULFID   723   737       BY SIMILARITY.
FT   DISULFID   739   748       BY SIMILARITY.
FT   DISULFID   755   766       BY SIMILARITY.
FT   DISULFID   760   775       BY SIMILARITY.
FT   DISULFID   777   786       BY SIMILARITY.
FT   DISULFID   793   804       BY SIMILARITY.
```

```
FT   DISULFID    798    813       BY SIMILARITY.
FT   DISULFID    815    824       BY SIMILARITY.
FT   DISULFID    831    842       BY SIMILARITY.
FT   DISULFID    836    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
FT   DISULFID    871    882       BY SIMILARITY.
FT   DISULFID    876    891       BY SIMILARITY.
FT   DISULFID    893    902       BY SIMILARITY.
FT   DISULFID    909    920       BY SIMILARITY.
FT   DISULFID    914    929       BY SIMILARITY.
FT   DISULFID    931    940       BY SIMILARITY.
FT   DISULFID    947    958       BY SIMILARITY.
FT   DISULFID    952    967       BY SIMILARITY.
FT   DISULFID    969    978       BY SIMILARITY.
FT   DISULFID   1023   1034       BY SIMILARITY.
FT   DISULFID   1028   1043       BY SIMILARITY.
FT   DISULFID   1045   1054       BY SIMILARITY.
FT   DISULFID   1061   1072       BY SIMILARITY.
FT   DISULFID   1066   1081       BY SIMILARITY.
FT   DISULFID   1083   1092       BY SIMILARITY.
FT   DISULFID   1099   1120       BY SIMILARITY.
FT   DISULFID   1114   1129       BY SIMILARITY.
FT   DISULFID   1131   1140       BY SIMILARITY.
FT   DISULFID   1147   1158       BY SIMILARITY.
FT   DISULFID   1152   1167       BY SIMILARITY.
FT   DISULFID   1169   1178       BY SIMILARITY.
FT   DISULFID   1185   1196       BY SIMILARITY.
FT   DISULFID   1190   1205       BY SIMILARITY.
FT   DISULFID   1207   1216       BY SIMILARITY.
FT   DISULFID   1223   1242       BY SIMILARITY.
FT   DISULFID   1236   1251       BY SIMILARITY.
FT   DISULFID   1253   1262       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           13.9%;  Score 113;  DB 1;  Length 2437;
  Best Local Similarity  51.6%;  Pred. No. 2.92e-07;
  Matches   16;  Conservative    5;  Mismatches   7;  Indels   3;  Gaps   3;

Db   1361 C-VSGHLSPRCLCAPGFSGHECQTRMDSPCL 1390
         | :| :  | ||| |||||  || : ::|||
Qy      2 CHISDQGEPYCLCQPGFSGEHCQ-Q-ENPCL 30


RESULT    8
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643       EXTRACELLULAR.
FT   TRANSMEM  1644   1664       POTENTIAL.
FT   DOMAIN    1665   2318       CYTOPLASMIC.
FT   DOMAIN      39   1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261       PEST.
FT   DOMAIN      39     78       EGF-LIKE 1.
FT   DOMAIN      79    119       EGF-LIKE 2.
FT   DOMAIN     120    157       EGF-LIKE 3.
FT   DOMAIN     159    196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235       EGF-LIKE 5.
FT   DOMAIN     237    273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313       EGF-LIKE 7.
FT   DOMAIN     315    351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390       EGF-LIKE 9.
FT   DOMAIN     392    430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731       EGF-LIKE 18.
FT   DOMAIN     735    771       EGF-LIKE 19.
FT   DOMAIN     772    809       EGF-LIKE 20.
FT   DOMAIN     811    848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961       EGF-LIKE 24.
FT   DOMAIN     963    999       EGF-LIKE 25.
FT   DOMAIN    1001   1035       EGF-LIKE 26.
FT   DOMAIN    1037   1083       EGF-LIKE 27.
FT   DOMAIN    1085   1121       EGF-LIKE 28.
FT   DOMAIN    1123   1159       EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245       EGF-LIKE 31.
FT   DOMAIN    1247   1288       EGF-LIKE 32.
FT   DOMAIN    1290   1326       EGF-LIKE 33.
FT   DOMAIN    1336   1374       EGF-LIKE 34.
FT   REPEAT    1388   1428       LIN/NOTCH 1.
FT   REPEAT    1429   1467       LIN/NOTCH 2.
FT   REPEAT    1468   1503       LIN/NOTCH 3.
FT   REPEAT    1784   1816       CDC10/SWI6 1.
FT   REPEAT    1817   1865       CDC10/SWI6 2.
FT   REPEAT    1866   1898       CDC10/SWI6 3.
FT   REPEAT    1899   1932       CDC10/SWI6 4.
FT   REPEAT    1933   1965       CDC10/SWI6 5.
FT   REPEAT    1966   1998       CDC10/SWI6 6.
FT   DISULFID    43     55       BY SIMILARITY.
FT   DISULFID    49     66       BY SIMILARITY.
FT   DISULFID    68     77       BY SIMILARITY.
```

```
FT   DISULFID    83     94       BY SIMILARITY.
FT   DISULFID    88    107       BY SIMILARITY.
FT   DISULFID   109    118       BY SIMILARITY.
FT   DISULFID   124    135       BY SIMILARITY.
FT   DISULFID   129    145       BY SIMILARITY.
FT   DISULFID   147    156       BY SIMILARITY.
FT   DISULFID   163    175       BY SIMILARITY.
FT   DISULFID   169    184       BY SIMILARITY.
FT   DISULFID   186    195       BY SIMILARITY.
FT   DISULFID   202    213       BY SIMILARITY.
FT   DISULFID   207    223       BY SIMILARITY.
FT   DISULFID   225    234       BY SIMILARITY.
FT   DISULFID   241    252       BY SIMILARITY.
FT   DISULFID   246    261       BY SIMILARITY.
FT   DISULFID   263    272       BY SIMILARITY.
FT   DISULFID   279    292       BY SIMILARITY.
FT   DISULFID   286    301       BY SIMILARITY.
FT   DISULFID   303    312       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   324    339       BY SIMILARITY.
FT   DISULFID   341    350       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    378       BY SIMILARITY.
FT   DISULFID   380    389       BY SIMILARITY.
FT   DISULFID   396    409       BY SIMILARITY.
FT   DISULFID   403    418       BY SIMILARITY.
FT   DISULFID   420    429       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   441    456       BY SIMILARITY.
FT   DISULFID   458    467       BY SIMILARITY.
FT   DISULFID   474    485       BY SIMILARITY.
FT   DISULFID   479    494       BY SIMILARITY.
FT   DISULFID   496    505       BY SIMILARITY.
FT   DISULFID   512    523       BY SIMILARITY.
FT   DISULFID   517    532       BY SIMILARITY.
FT   DISULFID   534    543       BY SIMILARITY.
FT   DISULFID   550    560       BY SIMILARITY.
FT   DISULFID   555    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   587    598       BY SIMILARITY.
FT   DISULFID   592    607       BY SIMILARITY.
FT   DISULFID   609    618       BY SIMILARITY.
FT   DISULFID   625    635       BY SIMILARITY.
FT   DISULFID   630    644       BY SIMILARITY.
FT   DISULFID   646    655       BY SIMILARITY.
FT   DISULFID   662    673       BY SIMILARITY.
FT   DISULFID   667    682       BY SIMILARITY.
FT   DISULFID   684    693       BY SIMILARITY.
FT   DISULFID   700    710       BY SIMILARITY.
FT   DISULFID   705    719       BY SIMILARITY.
FT   DISULFID   721    730       BY SIMILARITY.
FT   DISULFID   739    750       BY SIMILARITY.
FT   DISULFID   744    759       BY SIMILARITY.
FT   DISULFID   761    770       BY SIMILARITY.
FT   DISULFID   776    787       BY SIMILARITY.
FT   DISULFID   781    797       BY SIMILARITY.
FT   DISULFID   799    808       BY SIMILARITY.
FT   DISULFID   815    827       BY SIMILARITY.
FT   DISULFID   821    836       BY SIMILARITY.
FT   DISULFID   838    847       BY SIMILARITY.
FT   DISULFID   854    865       BY SIMILARITY.
FT   DISULFID   859    874       BY SIMILARITY.
FT   DISULFID   876    885       BY SIMILARITY.
FT   DISULFID   892    902       BY SIMILARITY.
FT   DISULFID   897    911       BY SIMILARITY.
FT   DISULFID   913    922       BY SIMILARITY.
FT   DISULFID   929    940       BY SIMILARITY.
FT   DISULFID   934    949       BY SIMILARITY.
FT   DISULFID   951    960       BY SIMILARITY.
FT   DISULFID   967    978       BY SIMILARITY.
FT   DISULFID   972    987       BY SIMILARITY.
FT   DISULFID   989    998       BY SIMILARITY.
FT   DISULFID  1005   1016       BY SIMILARITY.
FT   DISULFID  1010   1023       BY SIMILARITY.
FT   DISULFID  1025   1034       BY SIMILARITY.
FT   DISULFID  1041   1062       BY SIMILARITY.
FT   DISULFID  1056   1071       BY SIMILARITY.
FT   DISULFID  1073   1082       BY SIMILARITY.
FT   DISULFID  1089   1100       BY SIMILARITY.
FT   DISULFID  1094   1109       BY SIMILARITY.
FT   DISULFID  1111   1120       BY SIMILARITY.
FT   DISULFID  1127   1138       BY SIMILARITY.
FT   DISULFID  1132   1147       BY SIMILARITY.
FT   DISULFID  1149   1158       BY SIMILARITY.
FT   DISULFID  1165   1183       BY SIMILARITY.
FT   DISULFID  1177   1192       BY SIMILARITY.
FT   DISULFID  1194   1203       BY SIMILARITY.
FT   DISULFID  1210   1223       BY SIMILARITY.
FT   DISULFID  1215   1233       BY SIMILARITY.
FT   DISULFID  1235   1244       BY SIMILARITY.
FT   DISULFID  1251   1262       BY SIMILARITY.
FT   DISULFID  1256   1276       BY SIMILARITY.
FT   DISULFID  1278   1287       BY SIMILARITY.
FT   DISULFID  1294   1305       BY SIMILARITY.
FT   DISULFID  1299   1314       BY SIMILARITY.
FT   DISULFID  1316   1325       BY SIMILARITY.
FT   DISULFID  1340   1351       BY SIMILARITY.
FT   DISULFID  1345   1362       BY SIMILARITY.
FT   DISULFID  1364   1373       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          13.8%;  Score 112;  DB 1;  Length 2318;
  Best Local Similarity  50.0%;  Pred. No. 4.65e-07;
  Matches   11;  Conservative    8;  Mismatches   2;  Indels   1;  Gaps   1;

Db    949 CLCRPGYTGTHCQYEADPCFSR 970
          |||:||::| ||| : :||:::
Qy     12 CLCQPGFSGEHCQ-QENPCLGQ 32


RESULT    9
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
FT   DOMAIN     869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907    943       EGF-LIKE 24.
FT   DOMAIN     945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983   1019       EGF-LIKE 26.
FT   DOMAIN    1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059   1095       EGF-LIKE 28.
FT   DOMAIN    1097   1143       EGF-LIKE 29.
FT   DOMAIN    1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267   1305       EGF-LIKE 33.
FT   DOMAIN    1307   1346       EGF-LIKE 34.
FT   DOMAIN    1348   1384       EGF-LIKE 35.
FT   DOMAIN    1387   1426       EGF-LIKE 36.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1865   2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2076       ANK MOTIF 6.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
FT   DISULFID   339    350       BY SIMILARITY.
FT   DISULFID   344    359       BY SIMILARITY.
FT   DISULFID   361    370       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    398       BY SIMILARITY.
FT   DISULFID   400    409       BY SIMILARITY.
FT   DISULFID   416    429       BY SIMILARITY.
FT   DISULFID   423    438       BY SIMILARITY.
FT   DISULFID   440    449       BY SIMILARITY.
FT   DISULFID   456    467       BY SIMILARITY.
FT   DISULFID   461    476       BY SIMILARITY.
FT   DISULFID   478    487       BY SIMILARITY.
FT   DISULFID   494    505       BY SIMILARITY.
FT   DISULFID   499    514       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   532    543       BY SIMILARITY.
FT   DISULFID   537    552       BY SIMILARITY.
FT   DISULFID   554    563       BY SIMILARITY.
FT   DISULFID   570    580       BY SIMILARITY.
FT   DISULFID   575    589       BY SIMILARITY.
FT   DISULFID   591    600       BY SIMILARITY.
FT   DISULFID   607    618       BY SIMILARITY.
FT   DISULFID   612    627       BY SIMILARITY.
FT   DISULFID   629    638       BY SIMILARITY.
FT   DISULFID   645    655       BY SIMILARITY.
FT   DISULFID   650    664       BY SIMILARITY.
FT   DISULFID   666    675       BY SIMILARITY.
FT   DISULFID   682    693       BY SIMILARITY.
FT   DISULFID   687    702       BY SIMILARITY.
FT   DISULFID   704    713       BY SIMILARITY.
FT   DISULFID   720    730       BY SIMILARITY.
FT   DISULFID   725    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    855       BY SIMILARITY.
FT   DISULFID   857    866       BY SIMILARITY.
FT   DISULFID   873    884       BY SIMILARITY.
FT   DISULFID   878    893       BY SIMILARITY.
FT   DISULFID   895    904       BY SIMILARITY.
FT   DISULFID   911    922       BY SIMILARITY.
FT   DISULFID   916    931       BY SIMILARITY.
FT   DISULFID   933    942       BY SIMILARITY.
FT   DISULFID   987    998       BY SIMILARITY.
FT   DISULFID   992   1007       BY SIMILARITY.
FT   DISULFID  1009   1018       BY SIMILARITY.
FT   DISULFID  1025   1036       BY SIMILARITY.
FT   DISULFID  1030   1045       BY SIMILARITY.
FT   DISULFID  1047   1056       BY SIMILARITY.
FT   DISULFID  1063   1074       BY SIMILARITY.
FT   DISULFID  1068   1083       BY SIMILARITY.
FT   DISULFID  1085   1094       BY SIMILARITY.
FT   DISULFID  1101   1122       BY SIMILARITY.
FT   DISULFID  1116   1131       BY SIMILARITY.
```

```
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.
FT   DISULFID   1225   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1271   1284       BY SIMILARITY.
FT   DISULFID   1276   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1311   1322       BY SIMILARITY.
FT   DISULFID   1316   1334       BY SIMILARITY.
FT   DISULFID   1336   1345       BY SIMILARITY.
FT   DISULFID   1352   1363       BY SIMILARITY.
FT   DISULFID   1357   1372       BY SIMILARITY.
FT   DISULFID   1374   1383       BY SIMILARITY.
FT   DISULFID   1391   1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          13.7%;  Score 111;  DB 1;  Length 2531;
  Best Local Similarity  35.5%;  Pred. No. 7.40e-07;
  Matches   11;  Conservative   12;  Mismatches   7;  Indels   1;  Gaps   1;

Db     36 RCEVANGTEA-CVCSGAFVGQRCQDPSPCLS 65
          :| :::  |: |:|  :| |::||: :|||:
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLG 31


RESULT   10
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN       19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1726   1746       POTENTIAL.
FT   DOMAIN     1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN     1449   1462       CYS-RICH.
FT   DOMAIN     1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1445   1480       LIN/NOTCH 1.
FT   REPEAT     1481   1522       LIN/NOTCH 2.
FT   REPEAT     1523   1562       LIN/NOTCH 3.
FT   DOMAIN     1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT     1865   1910       ANK MOTIF 1.
FT   REPEAT     1912   1942       ANK MOTIF 2.
FT   REPEAT     1944   1975       ANK MOTIF 3.
FT   REPEAT     1978   2009       ANK MOTIF 4.
FT   REPEAT     2011   2042       ANK MOTIF 5.
FT   REPEAT     2044   2075       ANK MOTIF 6.
FT   CARBOHYD    888    888       POTENTIAL.
FT   CARBOHYD    959    959       POTENTIAL.
FT   CARBOHYD   1179   1179       POTENTIAL.
FT   CARBOHYD   1241   1241       POTENTIAL.
FT   CARBOHYD   1489   1489       POTENTIAL.
FT   CARBOHYD   1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match          13.7%;  Score 111;  DB 1;  Length 2531;
  Best Local Similarity  24.2%;  Pred. No. 7.40e-07;
  Matches   24;  Conservative   28;  Mismatches   39;  Indels   8;  Gaps   8;

Db    589 CLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNSYLCLCLKGTTGPNCEINLDDCASN 648
          ||||||::| ||: : | | :|  |     : |:::|   |  ::  | |     |::
Qy     12 CLCQPGFSGEHCQQE-NPCLGQVVREV-I-R-RQKGYAS-CATASKVPIMECR-GGCGPQ 65

Db    649 PCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            |::     :   ::|:  :::   : : :: ||:  :|
Qy     66 CCQPTRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCLAC 102


RESULT   11
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
```

```
RL   J. VIROL. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG, AND TESTIS;
RX   MEDLINE; 96281668.
RA   UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   DEVELOPMENT 122:2251-2259(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; G1714084; -.
DR   EMBL; U43691; G1401160; -.
DR   PIR; A38072; TVMVT3.
DR   MGD; MGI:107471; NOTCH4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   DOMAIN      21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1444   1464       POTENTIAL.
FT   DOMAIN    1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     60       EGF-LIKE 1.
FT   DOMAIN      61    112       EGF-LIKE 2.
FT   DOMAIN     115    152       EGF-LIKE 3.
FT   DOMAIN     153    189       EGF-LIKE 4.
FT   DOMAIN     191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     231    271       EGF-LIKE 6.
FT   DOMAIN     273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     389    427       EGF-LIKE 10.
FT   DOMAIN     429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     622    656       EGF-LIKE 16.
FT   DOMAIN     658    686       EGF-LIKE 17.
FT   DOMAIN     688    724       EGF-LIKE 18.
FT   DOMAIN     726    762       EGF-LIKE 19.
FT   DOMAIN     764    800       EGF-LIKE 20.
FT   DOMAIN     803    839       EGF-LIKE 21.
FT   DOMAIN     841    877       EGF-LIKE 22.
FT   DOMAIN     878    924       EGF-LIKE 23.
FT   DOMAIN     926    962       EGF-LIKE 24.
FT   DOMAIN     964   1000       EGF-LIKE 25.
FT   DOMAIN    1002   1040       EGF-LIKE 26.
FT   DOMAIN    1042   1081       EGF-LIKE 27.
FT   DOMAIN    1083   1122       EGF-LIKE 28.
FT   DOMAIN    1126   1167       EGF-LIKE 29.
FT   DOMAIN    1168   1282       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1168   1208       LIN/NOTCH 1.
FT   REPEAT    1209   1242       LIN/NOTCH 2.
FT   REPEAT    1243   1282       LIN/NOTCH 3.
FT   DOMAIN    1572   1785       6 X ANK MOTIF REPEATS.
FT   REPEAT    1572   1603       ANK MOTIF 1.
FT   REPEAT    1622   1653       ANK MOTIF 2.
FT   REPEAT    1654   1685       ANK MOTIF 3.
FT   REPEAT    1688   1719       ANK MOTIF 4.
FT   REPEAT    1721   1752       ANK MOTIF 5.
FT   REPEAT    1754   1785       ANK MOTIF 6.
FT   DISULFID    25     38       BY SIMILARITY.
FT   DISULFID    32     48       BY SIMILARITY.
FT   DISULFID    50     59       BY SIMILARITY.
FT   DISULFID    65     77       BY SIMILARITY.
FT   DISULFID    71    100       BY SIMILARITY.
FT   DISULFID   102    111       BY SIMILARITY.
FT   DISULFID   119    130       BY SIMILARITY.
FT   DISULFID   124    140       BY SIMILARITY.
FT   DISULFID   142    151       BY SIMILARITY.
FT   DISULFID   157    168       BY SIMILARITY.
FT   DISULFID   162    177       BY SIMILARITY.
FT   DISULFID   179    188       BY SIMILARITY.
FT   DISULFID   195    208       BY SIMILARITY.
FT   DISULFID   202    217       BY SIMILARITY.
FT   DISULFID   219    228       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   277    288       BY SIMILARITY.
FT   DISULFID   282    297       BY SIMILARITY.
FT   DISULFID   299    308       BY SIMILARITY.
FT   DISULFID   315    329       BY SIMILARITY.
FT   DISULFID   323    338       BY SIMILARITY.
FT   DISULFID   340    349       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    376       BY SIMILARITY.
FT   DISULFID   378    387       BY SIMILARITY.
FT   DISULFID   393    404       BY SIMILARITY.
FT   DISULFID   398    415       BY SIMILARITY.
FT   DISULFID   417    426       BY SIMILARITY.
FT   DISULFID   433    449       BY SIMILARITY.
FT   DISULFID   443    458       BY SIMILARITY.
FT   DISULFID   460    469       BY SIMILARITY.
FT   DISULFID   476    487       BY SIMILARITY.
FT   DISULFID   481    496       BY SIMILARITY.
FT   DISULFID   498    507       BY SIMILARITY.
FT   DISULFID   514    525       BY SIMILARITY.
FT   DISULFID   519    534       BY SIMILARITY.
FT   DISULFID   536    545       BY SIMILARITY.
FT   DISULFID   552    563       BY SIMILARITY.
FT   DISULFID   557    572       BY SIMILARITY.
FT   DISULFID   574    583       BY SIMILARITY.
FT   DISULFID   590    601       BY SIMILARITY.
FT   DISULFID   595    610       BY SIMILARITY.
FT   DISULFID   612    621       BY SIMILARITY.
FT   DISULFID   626    637       BY SIMILARITY.
FT   DISULFID   631    646       BY SIMILARITY.
FT   DISULFID   648    655       BY SIMILARITY.
FT   DISULFID   662    669       BY SIMILARITY.
FT   DISULFID   664    674       BY SIMILARITY.
FT   DISULFID   676    685       BY SIMILARITY.
FT   DISULFID   692    703       BY SIMILARITY.
FT   DISULFID   697    712       BY SIMILARITY.
FT   DISULFID   714    723       BY SIMILARITY.
FT   DISULFID   730    741       BY SIMILARITY.
FT   DISULFID   735    750       BY SIMILARITY.
```

```
FT   DISULFID   752   761       BY SIMILARITY.
FT   DISULFID   768   779       BY SIMILARITY.
FT   DISULFID   773   788       BY SIMILARITY.
FT   DISULFID   790   799       BY SIMILARITY.
FT   DISULFID   807   818       BY SIMILARITY.
FT   DISULFID   812   827       BY SIMILARITY.
FT   DISULFID   829   838       BY SIMILARITY.
FT   DISULFID   845   856       BY SIMILARITY.
FT   DISULFID   850   865       BY SIMILARITY.
FT   DISULFID   867   876       BY SIMILARITY.
FT   DISULFID   882   903       BY SIMILARITY.
FT   DISULFID   897   912       BY SIMILARITY.
FT   DISULFID   914   923       BY SIMILARITY.
FT   DISULFID   930   941       BY SIMILARITY.
FT   DISULFID   935   950       BY SIMILARITY.
FT   DISULFID   952   961       BY SIMILARITY.
FT   DISULFID   968   979       BY SIMILARITY.
FT   DISULFID   973   988       BY SIMILARITY.
FT   DISULFID   990   999       BY SIMILARITY.
FT   DISULFID  1006  1019       BY SIMILARITY.
FT   DISULFID  1011  1028       BY SIMILARITY.
FT   DISULFID  1030  1039       BY SIMILARITY.
FT   DISULFID  1046  1057       BY SIMILARITY.
FT   DISULFID  1051  1069       BY SIMILARITY.
FT   DISULFID  1071  1080       BY SIMILARITY.
FT   DISULFID  1087  1098       BY SIMILARITY.
FT   DISULFID  1092  1110       BY SIMILARITY.
FT   DISULFID  1112  1121       BY SIMILARITY.
FT   DISULFID  1130  1142       BY SIMILARITY.
FT   DISULFID  1136  1155       BY SIMILARITY.
FT   DISULFID  1157  1166       BY SIMILARITY.
FT   CARBOHYD   711   711       POTENTIAL.
FT   CARBOHYD   960   960       POTENTIAL.
FT   CARBOHYD  1139  1139       POTENTIAL.
FT   CONFLICT    43    43       Q -> R (IN REF. 3).
FT   CONFLICT   298   298       L -> P (IN REF. 3).
FT   CONFLICT   884   884       M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match          13.5%;  Score 110;  DB 1;  Length 1964;
  Best Local Similarity  43.8%;  Pred. No. 1.18e-06;
  Matches   14;  Conservative    7;  Mismatches   9;  Indels   2;  Gaps   2;

Db    449 C-INTPGSFNCLCLPGYTGSRCEADHNECLSQ 479
          | |:  |   ||| ||::| :|: : | ||:|
        2 CHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32


RESULT  12
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    19       POTENTIAL.
FT   CHAIN       20  2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20  1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729  1750       POTENTIAL.
FT   DOMAIN    1751  2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    57       EGF-LIKE 1.
FT   DOMAIN      58    99       EGF-LIKE 2.
FT   DOMAIN     102   140       EGF-LIKE 3.
FT   DOMAIN     141   177       EGF-LIKE 4.
FT   DOMAIN     179   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217   254       EGF-LIKE 6.
FT   DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371   409       EGF-LIKE 10.
FT   DOMAIN     411   449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451   487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489   525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527   563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565   600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602   638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     640   675       EGF-LIKE 17.
FT   DOMAIN     677   713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715   750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     752   788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     790   826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     828   866       EGF-LIKE 22.
FT   DOMAIN     868   904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     906   942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     944   980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     982  1018       EGF-LIKE 26.
FT   DOMAIN    1020  1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1058  1094       EGF-LIKE 28.
FT   DOMAIN    1096  1142       EGF-LIKE 29.
FT   DOMAIN    1144  1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1182  1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1220  1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1266  1304       EGF-LIKE 33.
FT   DOMAIN    1306  1346       EGF-LIKE 34.
FT   DOMAIN    1347  1383       EGF-LIKE 35.
FT   DOMAIN    1386  1424       EGF-LIKE 36.
FT   DOMAIN    1441  1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1441  1478       LIN/NOTCH 1.
FT   REPEAT    1479  1520       LIN/NOTCH 2.
FT   REPEAT    1521  1560       LIN/NOTCH 3.
FT   DOMAIN    1871  2083       6 X ANK MOTIF REPEATS.
FT   DISULFID    22    35       BY SIMILARITY.
FT   DISULFID    29    45       BY SIMILARITY.
FT   DISULFID    47    56       BY SIMILARITY.
```

```
FT   DISULFID    62    74       BY SIMILARITY.
FT   DISULFID    68    87       BY SIMILARITY.
FT   DISULFID    89    98       BY SIMILARITY.
FT   DISULFID   106   117       BY SIMILARITY.
FT   DISULFID   111   128       BY SIMILARITY.
FT   DISULFID   130   139       BY SIMILARITY.
FT   DISULFID   145   156       BY SIMILARITY.
FT   DISULFID   150   165       BY SIMILARITY.
FT   DISULFID   167   176       BY SIMILARITY.
FT   DISULFID   183   194       BY SIMILARITY.
FT   DISULFID   188   203       BY SIMILARITY.
FT   DISULFID   205   214       BY SIMILARITY.
FT   DISULFID   221   232       BY SIMILARITY.
FT   DISULFID   226   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   265   280       BY SIMILARITY.
FT   DISULFID   282   291       BY SIMILARITY.
FT   DISULFID   298   311       BY SIMILARITY.
FT   DISULFID   305   320       BY SIMILARITY.
FT   DISULFID   322   331       BY SIMILARITY.
FT   DISULFID   338   349       BY SIMILARITY.
FT   DISULFID   343   358       BY SIMILARITY.
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   375   386       BY SIMILARITY.
FT   DISULFID   380   397       BY SIMILARITY.
FT   DISULFID   399   408       BY SIMILARITY.
FT   DISULFID   415   428       BY SIMILARITY.
FT   DISULFID   422   437       BY SIMILARITY.
FT   DISULFID   439   448       BY SIMILARITY.
FT   DISULFID   455   466       BY SIMILARITY.
FT   DISULFID   460   475       BY SIMILARITY.
FT   DISULFID   477   486       BY SIMILARITY.
FT   DISULFID   493   504       BY SIMILARITY.
FT   DISULFID   498   513       BY SIMILARITY.
FT   DISULFID   515   524       BY SIMILARITY.
FT   DISULFID   531   542       BY SIMILARITY.
FT   DISULFID   536   551       BY SIMILARITY.
FT   DISULFID   553   562       BY SIMILARITY.
FT   DISULFID   569   579       BY SIMILARITY.
FT   DISULFID   574   588       BY SIMILARITY.
FT   DISULFID   590   599       BY SIMILARITY.
FT   DISULFID   606   617       BY SIMILARITY.
FT   DISULFID   611   626       BY SIMILARITY.
FT   DISULFID   628   637       BY SIMILARITY.
FT   DISULFID   644   654       BY SIMILARITY.
FT   DISULFID   649   663       BY SIMILARITY.
FT   DISULFID   665   674       BY SIMILARITY.
FT   DISULFID   681   692       BY SIMILARITY.
FT   DISULFID   686   701       BY SIMILARITY.
FT   DISULFID   703   712       BY SIMILARITY.
FT   DISULFID   719   729       BY SIMILARITY.
FT   DISULFID   724   738       BY SIMILARITY.
FT   DISULFID   740   749       BY SIMILARITY.
FT   DISULFID   756   767       BY SIMILARITY.
FT   DISULFID   761   776       BY SIMILARITY.
FT   DISULFID   778   787       BY SIMILARITY.
FT   DISULFID   794   805       BY SIMILARITY.
FT   DISULFID   799   814       BY SIMILARITY.
FT   DISULFID   816   825       BY SIMILARITY.
FT   DISULFID   832   843       BY SIMILARITY.
FT   DISULFID   837   854       BY SIMILARITY.
FT   DISULFID   856   865       BY SIMILARITY.
FT   DISULFID   872   883       BY SIMILARITY.
FT   DISULFID   877   892       BY SIMILARITY.
FT   DISULFID   894   903       BY SIMILARITY.
FT   DISULFID   910   921       BY SIMILARITY.
FT   DISULFID   915   930       BY SIMILARITY.
FT   DISULFID   932   941       BY SIMILARITY.
FT   DISULFID   986   997       BY SIMILARITY.
FT   DISULFID   991  1006       BY SIMILARITY.
FT   DISULFID  1008  1017       BY SIMILARITY.
FT   DISULFID  1024  1035       BY SIMILARITY.
FT   DISULFID  1029  1044       BY SIMILARITY.
FT   DISULFID  1046  1055       BY SIMILARITY.
FT   DISULFID  1062  1073       BY SIMILARITY.
FT   DISULFID  1067  1082       BY SIMILARITY.
FT   DISULFID  1084  1093       BY SIMILARITY.
FT   DISULFID  1100  1121       BY SIMILARITY.
FT   DISULFID  1115  1130       BY SIMILARITY.
FT   DISULFID  1132  1141       BY SIMILARITY.
FT   DISULFID  1148  1159       BY SIMILARITY.
FT   DISULFID  1153  1168       BY SIMILARITY.
FT   DISULFID  1170  1179       BY SIMILARITY.
FT   DISULFID  1186  1197       BY SIMILARITY.
FT   DISULFID  1191  1206       BY SIMILARITY.
FT   DISULFID  1208  1217       BY SIMILARITY.
FT   DISULFID  1224  1243       BY SIMILARITY.
FT   DISULFID  1237  1252       BY SIMILARITY.
FT   DISULFID  1254  1263       BY SIMILARITY.
FT   DISULFID  1270  1283       BY SIMILARITY.
FT   DISULFID  1275  1292       BY SIMILARITY.
FT   DISULFID  1294  1303       BY SIMILARITY.
FT   DISULFID  1310  1321       BY SIMILARITY.
FT   DISULFID  1315  1333       BY SIMILARITY.
FT   DISULFID  1335  1344       BY SIMILARITY.
FT   DISULFID  1351  1362       BY SIMILARITY.
FT   DISULFID  1356  1371       BY SIMILARITY.
FT   DISULFID  1373  1382       BY SIMILARITY.
FT   DISULFID  1390  1401       BY SIMILARITY.
FT   DISULFID  1395  1412       BY SIMILARITY.
FT   DISULFID  1414  1423       BY SIMILARITY.
FT   CARBOHYD   462   462       POTENTIAL.
FT   CARBOHYD   887   887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          13.3%;  Score 108;  DB 1;  Length 2524;
  Best Local Similarity  22.2%;  Pred. No. 2.95e-06;
  Matches   22;  Conservative   30;  Mismatches  39;  Indels   8;  Gaps   7;

Db    588 CLCRPGYTGRLCDNDINECLSKPCLNGGQCTDRENGYICTCPKGTTGVNCETKIDDCASN 647
          |||:||::|  |: : | ||:   :   :   |::||  :|: ::     | :   |::
Qy     12 CLCQPGFSGEHCQQE-NPCLGQV-VR--EVIRRQKGYA-SCATASKVPIMECR-GGCGPQ 65

Db    648 LCDNGKCIDKIDGYECTCEPGYTGKLCNININECDSNPC 686
            |:  :   :   ::||  :::   : : :: ||   :|
Qy     66 CCQPTRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCLAC 102


RESULT   13
ID   VWF_HUMAN      STANDARD;      PRT;  2813 AA.
AC   P04275;
DT   20-MAR-1987 (REL. 04, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   VON WILLEBRAND FACTOR PRECURSOR.
GN   F8VWF OR VWF.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90062044.
RA   MANCUSO D.J., TULEY E.A., WESTFIELD L.A., WORRALL N.K.,
RA   SHELTON-INLOES B.B., SORACE J.M., ALEVY Y.G., SADLER J.E.;
RT   "Structure of the gene for human von Willebrand factor.";
RL   J. BIOL. CHEM. 264:19514-19527(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87016349.
RA   BONTHRON D., ORR E.C., MITSOCK L.M., MITSOCK L.M., GINSBURG D.,
RA   HANDIN R.I., ORKIN S.H.;
RT   "Nucleotide sequence of pre-pro-von Willebrand factor cDNA.";
RL   NUCLEIC ACIDS RES. 14:7125-7128(1986).
RN   [3]
```

RP SEQUENCE OF 1-1400 FROM N.A.
RX MEDLINE; 87004550.
RA VERWEIJ C.L., DIERGAARDE P.J., HART M., PANNEKOEK H.;
RT "Full-length von Willebrand factor (vWF) cDNA encodes a highly
RT repetitive protein considerably larger than the mature vWF subunit.";
RL EMBO J. 5:1839-1847(1986).
RN [4]
RP ERRATUM.
RA VERWEIJ C.L., DIERGAARDE P.J., HART M., PANNEKOEK H.;
RL EMBO J. 5:3074-3074(1986).
RN [5]
RP SEQUENCE OF 764-2813.
RX MEDLINE; 86269895.
RA TITANI K., KUMAR S., TAKIO K., ERICSSON L.H., WADE R.D., ASHIDA K.,
RA WALSH K.A., CHOPEK M.W., SADLER J.E., FUJIKAWA K.;
RT "Amino acid sequence of human von Willebrand factor.";
RL BIOCHEMISTRY 25:3171-3184(1986).
RN [6]
RP SEQUENCE OF 781-1424 FROM N.A.
RX MEDLINE; 86269894.
RA SHELTON-INLOES B.B., TITANI K., SADLER J.E.;
RT "cDNA sequences for human von Willebrand factor reveal five types of
RT repeated domains and five possible protein sequence polymorphisms.";
RL BIOCHEMISTRY 25:3164-3171(1986).
RN [7]
RP SEQUENCE OF 764-873 AND 1289-2813 FROM N.A.
RX MEDLINE; 86016708.
RA SADLER J.E., SHELTON-INLOES B.B., SORACE J.M., HARLAN J.M.,
RA TITANI K., DAVIE E.W.;
RT "Cloning and characterization of two cDNAs coding for human von
RT Willebrand factor.";
RL PROC. NATL. ACAD. SCI. U.S.A. 82:6394-6398(1985).
RN [8]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE; 85269603.
RA VERWEIJ C.L., DE VRIES C.J.M., DISTEL B., VAN ZONNEVELD A.-J.,
RA GEURTS VAN KESSEL A., VAN MOURIK J.A., PANNEKOEK H.;
RT "Construction of cDNA coding for human von Willebrand factor using
RT antibody probes for colony-screening and mapping of the chromosomal
RT gene.";
RL NUCLEIC ACIDS RES. 13:4699-4717(1985).
RN [9]
RP SEQUENCE OF 1-177 FROM N.A.
RX MEDLINE; 88111704.
RA BONTHRON D., ORKIN S.H.;
RT "The human von Willebrand factor gene. Structure of the 5' region.";
RL EUR. J. BIOCHEM. 171:51-57(1988).
RN [10]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE; 87260814.
RA COLLINS C.J., UNDERDAHL J.P., LEVENE R.B., RAVERA C.P.,
RA MORIN M.J., DOMBALAGIAN M.J., RICCA G., LIVINGSTON D.M.,
RA LYNCH D.C.;
RT "Molecular cloning of the human gene for von Willebrand factor and
RT identification of the transcription initiation site.";
RL PROC. NATL. ACAD. SCI. U.S.A. 84:4393-4397(1987).
RN [11]
RP DISULFIDE BONDS.
RX MEDLINE; 88163465.
RA MARTI T., ROSSELET S.J., TITANI K., WALSH K.A.;
RT "Identification of disulfide-bridged substructures within human von
RT Willebrand factor.";
RL BIOCHEMISTRY 26:8099-8109(1987).
RN [12]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE; 86274702.
RA SAMOR B., MICHALSKI J.C., DEBRAY H., MAZURIER C., GOUDEMAND M.,
RA VAN HALBEEK H., VLIEGENTHART J.F.G., MONTREUIL J.;
RT "Primary structure of a new tetraantennary glycan of the N-
RT acetyllactosaminic type isolated from human factor VIII/von
RT Willebrand factor.";
RL EUR. J. BIOCHEM. 158:295-298(1986).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1685-1873.
RX MEDLINE; 97472999.
RA HUIZINGA E.G., MARTIJN VAN DER PLAS R., KROON J., SIXMA J.J., GROS P.;
RT "Crystal structure of the A3 domain of human von Willebrand factor:
RT implications for collagen binding.";
RL STRUCTURE 5:1147-1156(1997).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1686-1872.
RX MEDLINE; 97460108.
RA BIENKOWSKA J., CRUZ M., ATIEMO A., HANDIN R., LIDDINGTON R.;
RT "The von willebrand factor A3 domain does not contain a metal ion-
RT dependent adhesion site motif.";
RL J. BIOL. CHEM. 272:25162-25167(1997).
RN [15]
RP VARIANTS TRP-1597 AND ASP-1607.
RX MEDLINE; 89264495.
RA GINSBURG D., KONKLE B.A., GILL J.C., MONTGOMERY R.R.,
RA BOCKENSTEDT P.L., JOHNSON T.A., YANG A.Y.;
RT "Molecular basis of human von Willebrand disease: analysis of
RT platelet von Willebrand factor mRNA.";
RL PROC. NATL. ACAD. SCI. U.S.A. 86:3723-3727(1989).
RN [16]
RP VARIANT THR-1628.
RX MEDLINE; 91196734.
RA IANNUZZI M.C., HIDAKA N., BOEHNKE M., BRUCK M.E., HANNA W.T.,
RA COLLINS F.S., GINSBURG D.;
RT "Analysis of the relationship of von Willebrand disease (vWD) and
RT hereditary hemorrhagic telangiectasia and identification of a
RT potential type IIA vWD mutation (IIe865 to Thr).";
RL AM. J. HUM. GENET. 48:757-763(1991).
RN [17]
RP VARIANTS NORMANDY-2 AND NORMANDY-3.
RX MEDLINE; 92001464.
RA GAUCHER C., MERCIER B., JORIEUX S., OUFKIR D., MAZURIER C.;
RT "Identification of two point mutations in the von Willebrand factor
RT gene of three families with the 'Normandy' variant of von Willebrand
RT disease.";
RL BR. J. HAEMATOL. 78:506-514(1991).
RN [18]
RP VARIANT CYS-1308.
RX MEDLINE; 92104315.
RA DONNER M., ANDERSSON A.-M., KRISTOFFERSSON A.-C., NILSSON I.M.,
RA DAHLBACK B., HOLMBERG L.;
RT "An Arg545-->Cys545 substitution mutation of the von Willebrand
RT factor in type IIB von Willebrand's disease.";
RL EUR. J. HAEMATOL. 47:342-345(1991).
RN [19]
RP VARIANTS TRP-1306; CYS-1308 AND PRO-1613.
RX MEDLINE; 91185601.
RA RANDI A.M., RABINOWITZ I., MANCUSO D.J., MANNUCCI P.M., SADLER J.E.;
RT "Molecular basis of von Willebrand disease type IIB. Candidate
RT mutations cluster in one disulfide loop between proposed platelet
RT glycoprotein Ib binding sequences.";
RL J. CLIN. INVEST. 87:1220-1226(1991).
RN [20]
RP VARIANTS TRP-1306; CYS-1308; MET-1316; GLN-1341 AND HIS-1399.
RX MEDLINE; 91185602.
RA COONEY K.A., NICHOLS W.C., BRUCK M.E., BAHOU W.F., SHAPIRO A.D.,
RA BOWIE E.J.W., GRALNICK H.R., GINSBURG D.;
RT "The molecular defect in type IIB von Willebrand disease.
RT Identification of four potential missense mutations within the
RT putative GpIb binding domain.";
RL J. CLIN. INVEST. 87:1227-1233(1991).
RN [21]
RP VARIANT CYS-1313.
RX MEDLINE; 91187908.
RA WARE J., DENT J.A., AZUMA H., SUGIMOTO M., KYRLE P.A., YOSHIOKA A.,
RA RUGGERI Z.M.;
RT "Identification of a point mutation in type IIB von Willebrand
RT disease illustrating the regulation of von Willebrand factor affinity
RT for the platelet membrane glycoprotein Ib-IX receptor.";
RL PROC. NATL. ACAD. SCI. U.S.A. 88:2946-2950(1991).
RN [22]

```
RP   VARIANT NORMANDY-1.
RX   MEDLINE; 91296824.
RA   TULEY E.A., GAUCHER C., JORIEUX S., WORRALL N.K., SADLER J.E.,
RA   MAZURIER C.;
RT   "Expression of von Willebrand factor 'Normandy': an autosomal
RT   mutation that mimics hemophilia A.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 88:6377-6381(1991).
RN   [23]
RP   VARIANT MET-1316.
RX   MEDLINE; 92109240.
RA   MURRAY E.W., GILES A.R., LILLICRAP D.;
RT   "Germ-line mosaicism for a valine-to-methionine substitution at
RT   residue 553 in the glycoprotein Ib-binding domain of von Willebrand
RT   factor, causing type IIB von Willebrand disease.";
RL   AM. J. HUM. GENET. 50:199-207(1992).
RN   [24]
RP   VARIANTS TRP-1306; MET-1316; THR-1628 AND SER-1648.
RX   MEDLINE; 93042596.
RA   PIETU G., RIBBA A.S., DE PAILLETTE L., CHEREL G., LAVERGNE J.M.,
RA   BAHNAK B.R., MEYER D.;
RT   "Molecular study of von Willebrand disease: identification of
     potential mutations in patients with type IIA and type IIB.";
     BLOOD COAGUL. FIBRINOLYSIS 3:415-421(1992).
RN   [25]
RP   VARIANTS TRP-1306; CYS-1308; LEU-1314 AND LEU-1318.
RX   MEDLINE; 93041230.
RA   DONNER M., KRISTOFFERSSON A.C., LENK H., SCHEIBEL E., DAHLBACK B.,
RA   NILSSON I.M., HOLMBERG L.;
RT   "Type IIB von Willebrand's disease: gene mutations and clinical
...
Note: remainder of annotations omitted.

  Query Match          12.9%;  Score 105;  DB 1;  Length 2813;
  Best Local Similarity  41.2%;  Pred. No. 1.16e-05;
  Matches   14;  Conservative   5;  Mismatches  15;  Indels   0;  Gaps   0;

Db   2773 CCSPTRTEPMQVALHCTNGSVVYHEVLNAMECKC 2806
          || |||:      ::||:||    ||   :|| |
Qy     66 CCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT   14
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
DE   PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
     HSPG2.
     MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE; 92078153.
RA   NOONAN D.M., FULLE A., VALENTE P., CAI S., HORIGAN E., SASAKI M.,
RA   YAMADA Y., HASSELL J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. BIOL. CHEM. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 89034110.
RA   NOONAN D.M., HORIGAN E.A., LEDBETTER S.R., VOGELI G., SASAKI M.,
RA   YAMADA Y., HASSELL J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. BIOL. CHEM. 263:16379-16387(1988).
CC   -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC       MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC       CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC       PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC       SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC       SUBSTRATE FOR CELLS.
CC   -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC       DIMERS OR STELLATE STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       15 C2-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77174; G200296; -.
DR   EMBL; J04054; G200253; -.
DR   EMBL; J04055; G200301; -.
DR   MGD; MGI:96257; HSPG2.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PFAM; PF00047; ig; 14.
DR   PFAM; PF00052; laminin_B; 3.
DR   PFAM; PF00053; laminin_EGF; 8.
DR   PFAM; PF00054; laminin_G; 3.
DR   PFAM; PF00057; ldl_recept_a; 4.
DR   HSSP; P01130; 1AJJ.
KW   SIGNAL; BASEMENT MEMBRANE; PROTEOGLYCAN; REPEAT; GLYCOPROTEIN;
KW   HEPARAN SULFATE; LAMININ EGF-LIKE DOMAIN; IMMUNOGLOBULIN FOLD;
KW   EXTRACELLULAR MATRIX; EGF-LIKE DOMAIN.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN      22    193       DOMAIN I (UNIQUE, CONTAINS 3 HS SIDE
FT                                CHAINS).
FT   DOMAIN     194    403       DOMAIN II (4 LDLRA REPEATS).
FT   DOMAIN     404    504       DOMAIN IIA (1 IGG-REPEAT).
FT   DOMAIN     507   1676       DOMAIN III (SIMILAR TO SHORT ARM OF
FT                                LAMININ A CHAIN).
FT   DOMAIN    1677   2980       DOMAIN IV (SIMILAR TO NEURAL CELL
FT                                ADHESION MOLECULE; 14 IGG REPEATS).
FT   DOMAIN    2981   3707       DOMAIN V (C-TERMINAL G-DOMAIN OF LAMININ
FT                                ALPHA CHAINS AND EGF).
FT   DOMAIN     194    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN     281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     404    504       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN     531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN     731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN     764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN     814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN     879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN     924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN     934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN    1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN    1159   1208       LAMININ EGF-LIKE 6.
```

```
FT   DOMAIN     1209   1265      LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324      LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334      LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529      LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562      LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612      LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670      LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771      IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     1772   1865      IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     1866   1954      IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN     1955   2049      IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN     2050   2148      IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN     2149   2244      IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN     2245   2343      IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN     2344   2436      IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN     2437   2532      IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     2533   2619      IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN     2620   2720      IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN     2721   2809      IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN     2810   2895      IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN     2896   2980      IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN     2981   3130      LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT   DOMAIN     3049   3241      EGF-LIKE 1.
FT   DOMAIN     3304   3495      EGF-LIKE 2.
FT   DOMAIN     3558   3705      LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT   SITE         65     67      HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73      HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78      HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617      MEDIATES MOTOR NEURON ATTACHMENT
FT                               (POTENTIAL).
FT   DISULFID    199    212      BY SIMILARITY.
FT   DISULFID    206    225      BY SIMILARITY.
FT   DISULFID    219    234      BY SIMILARITY.
FT   DISULFID    285    297      BY SIMILARITY.
FT   DISULFID    292    310      BY SIMILARITY.
FT   DISULFID    304    319      BY SIMILARITY.
FT   DISULFID    325    337      BY SIMILARITY.
FT   DISULFID    332    350      BY SIMILARITY.
FT   DISULFID    344    359      BY SIMILARITY.
FT   DISULFID    368    381      BY SIMILARITY.
FT   DISULFID    375    394      BY SIMILARITY.
FT   DISULFID    388    403      BY SIMILARITY.
FT   DISULFID    428    479      BY SIMILARITY.
FT   DISULFID    764    773      BY SIMILARITY.
FT   DISULFID    766    780      BY SIMILARITY.
FT   DISULFID    783    792      BY SIMILARITY.
FT   DISULFID    795    811      BY SIMILARITY.
FT   DISULFID    814    829      BY SIMILARITY.
FT   DISULFID    816    839      BY SIMILARITY.
FT   DISULFID    842    851      BY SIMILARITY.
FT   DISULFID    854    869      BY SIMILARITY.
FT   DISULFID   1159   1168      BY SIMILARITY.
FT   DISULFID   1161   1175      BY SIMILARITY.
FT   DISULFID   1178   1187      BY SIMILARITY.
FT   DISULFID   1190   1206      BY SIMILARITY.
FT   DISULFID   1209   1224      BY SIMILARITY.
FT   DISULFID   1211   1234      BY SIMILARITY.
FT   DISULFID   1237   1246      BY SIMILARITY.
FT   DISULFID   1249   1263      BY SIMILARITY.
FT   DISULFID   1275   1287      BY SIMILARITY.
FT   DISULFID   1277   1293      BY SIMILARITY.
FT   DISULFID   1295   1304      BY SIMILARITY.
FT   DISULFID   1307   1322      BY SIMILARITY.
FT   DISULFID   1563   1572      BY SIMILARITY.
FT   DISULFID   1565   1579      BY SIMILARITY.
FT   DISULFID   1582   1591      BY SIMILARITY.
FT   DISULFID   1594   1610      BY SIMILARITY.
FT   DISULFID   1613   1628      BY SIMILARITY.
FT   DISULFID   1615   1638      BY SIMILARITY.
FT   DISULFID   1641   1650      BY SIMILARITY.
FT   DISULFID   1653   1668      BY SIMILARITY.
FT   DISULFID   1792   1839      BY SIMILARITY.
FT   DISULFID   1886   1932      BY SIMILARITY.
FT   DISULFID   1976   2021      BY SIMILARITY.
FT   DISULFID   2073   2118      BY SIMILARITY.
FT   DISULFID   2170   2215      BY SIMILARITY.
FT   DISULFID   2268   2313      BY SIMILARITY.
FT   DISULFID   2365   2413      BY SIMILARITY.
FT   DISULFID   2456   2506      BY SIMILARITY.
FT   DISULFID   2554   2599      BY SIMILARITY.
FT   DISULFID   2641   2686      BY SIMILARITY.
FT   DISULFID   2831   2876      BY SIMILARITY.
FT   DISULFID   2917   2962      BY SIMILARITY.
FT   CARBOHYD     65     65      GLYCOSAMINOGLYCAN (POTENTIAL).
FT   CARBOHYD     71     71      GLYCOSAMINOGLYCAN (POTENTIAL).
FT   CARBOHYD     76     76      GLYCOSAMINOGLYCAN (POTENTIAL).
FT   CARBOHYD     89     89      POTENTIAL.
FT   CARBOHYD    358    358      POTENTIAL.
FT   CARBOHYD    554    554      POTENTIAL.
FT   CARBOHYD   1256   1256      POTENTIAL.
FT   CARBOHYD   1891   1891      POTENTIAL.
FT   CARBOHYD   2336   2336      POTENTIAL.
FT   CARBOHYD   2394   2394      POTENTIAL.
FT   CARBOHYD   2427   2427      POTENTIAL.
FT   CARBOHYD   2600   2600      POTENTIAL.
FT   CARBOHYD   3098   3098      POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          12.9%;  Score 105;  DB 1;  Length 3707;
  Best Local Similarity  63.2%;  Pred. No. 1.16e-05;
  Matches   12;  Conservative    3;  Mismatches   3;  Indels   1;  Gaps   1;

Db   3446 CLCQDGFKGDLCEHEENPC 3464
          |||| || |: |::| |||
Qy     12 CLCQPGFSGEHCQQE-NPC 29


RESULT  15
ID   LRP1_CHICK     STANDARD;      PRT;  4543 AA.
AC   P98157;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 1 PRECURSOR (LRP)
DE   (ALPHA-2-MACROGLOBULIN RECEPTOR) (A2MR).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=LIVER, AND OVARY;
RX   MEDLINE; 94103212.
RA   NIMPF J., STIFANI S., BILOUS P.T., SCHNEIDER W.J.;
RT   "The somatic cell-specific low density lipoprotein receptor-related
RT   protein of the chicken. Close kinship to mammalian low density
RT   lipoprotein receptor gene family members.";
RL   J. BIOL. CHEM. 269:212-219(1994).
CC   -!- FUNCTION: INVOLVED IN THE PLASMA CLEARANCE OF CHYLOMICRON REMNANTS
CC       AND ACTIVATED ALPHA 2-MACROGLOBULIN, AS WELL AS THE LOCAL
CC       METABOLISM OF COMPLEXES BETWEEN PLASMINOGEN ACTIVATORS AND THEIR
CC       ENDOGENOUS INHIBITORS. BINDS VITELLOGENIN, CALCIUM AND ALPHA 2-
CC       MACROGLOBULIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SOMATIC.
CC   -!- PTM: CLEAVED INTO A 85 KD MEMBRANE-SPANNING SUBUNIT (LRP-85) AND
CC       A 515 KD LARGE EXTRACELLULAR DOMAIN (LRP-515) THAT REMAINS NON-
CC       COVALENTLY ASSOCIATED.
CC   -!- ALTERNATIVE PRODUCTS: IN CLONE JN18, AN ASP IS REPLACED BY
CC       SER-GLU-ARG-GLN-ASP DUE TO ALTERNATIVE SPLICING OF EXON3.
CC   -!- SIMILARITY: CONTAINS 22 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 31 LDL-RECEPTOR CLASS A DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74904; G438007; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01209; LDLRA_1; 27.
DR   PROSITE; PS50068; LDLRA_2; 31.
DR   PFAM; PF00008; EGF; 14.
DR   PFAM; PF00057; ldl_recept_a; 31.
DR   PFAM; PF00058; ldl_recept_b; 33.
DR   HSSP; P01130; 1AJJ.
KW   RECEPTOR; TRANSMEMBRANE; REPEAT; ENDOCYTOSIS; GLYCOPROTEIN;
KW   SIGNAL; CALCIUM-BINDING; EGF-LIKE DOMAIN; COATED PITS;
KW   ALTERNATIVE SPLICING.
FT   SIGNAL       1     21      POTENTIAL.
FT   CHAIN       22   4543      LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
FT                              PROTEIN 1.
FT   DOMAIN      22   4419      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  4420   4443      POTENTIAL.
FT   DOMAIN    4444   4543      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      27     68      LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      72    112      LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     113    151      EGF-LIKE 1.
FT   DOMAIN     152    191      EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     476    522      EGF-LIKE 3.
FT   DOMAIN     801    841      EGF-LIKE 4.
FT   DOMAIN     850    890      LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     891    931      LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     932    971      LDL-RECEPTOR CLASS A 5.
FT   DOMAIN     972   1011      LDL-RECEPTOR CLASS A 6.
FT   DOMAIN    1011   1051      LDL-RECEPTOR CLASS A 7.
FT   DOMAIN    1058   1097      LDL-RECEPTOR CLASS A 8.
FT   DOMAIN    1100   1140      LDL-RECEPTOR CLASS A 9.
FT   DOMAIN    1141   1180      LDL-RECEPTOR CLASS A 10.
FT   DOMAIN    1181   1220      EGF-LIKE 5.
FT   DOMAIN    1221   1260      EGF-LIKE 6.
FT   DOMAIN    1534   1577      EGF-LIKE 7.
FT   DOMAIN    1842   1883      EGF-LIKE 8.
FT   DOMAIN    2151   2191      EGF-LIKE 9.
FT   DOMAIN    2472   2512      EGF-LIKE 10.
FT   DOMAIN    2516   2557      LDL-RECEPTOR CLASS A 11.
FT   DOMAIN    2558   2596      LDL-RECEPTOR CLASS A 12.
FT   DOMAIN    2597   2635      LDL-RECEPTOR CLASS A 13.
FT   DOMAIN    2636   2684      LDL-RECEPTOR CLASS A 14.
FT   DOMAIN    2688   2730      LDL-RECEPTOR CLASS A 15.
FT   DOMAIN    2730   2769      LDL-RECEPTOR CLASS A 16.
FT   DOMAIN    2770   2812      LDL-RECEPTOR CLASS A 17.
FT   DOMAIN    2814   2853      LDL-RECEPTOR CLASS A 18.
FT   DOMAIN    2854   2897      LDL-RECEPTOR CLASS A 19.
FT   DOMAIN    2900   2938      LDL-RECEPTOR CLASS A 20.
FT   DOMAIN    2939   2978      EGF-LIKE 11.
FT   DOMAIN    2979   3019      EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    3287   3328      EGF-LIKE 13.
FT   DOMAIN    3329   3368      LDL-RECEPTOR CLASS A 21.
FT   DOMAIN    3369   3407      LDL-RECEPTOR CLASS A 22.
FT   DOMAIN    3408   3447      LDL-RECEPTOR CLASS A 23.
FT   DOMAIN    3448   3488      LDL-RECEPTOR CLASS A 24.
FT   DOMAIN    3489   3530      LDL-RECEPTOR CLASS A 25.
FT   DOMAIN    3531   3569      LDL-RECEPTOR CLASS A 26.
FT   DOMAIN    3570   3608      LDL-RECEPTOR CLASS A 27.
FT   DOMAIN    3608   3646      LDL-RECEPTOR CLASS A 28.
FT   DOMAIN    3649   3689      LDL-RECEPTOR CLASS A 29.
FT   DOMAIN    3690   3730      LDL-RECEPTOR CLASS A 30.
FT   DOMAIN    3736   3776      LDL-RECEPTOR CLASS A 31.
FT   DOMAIN    3779   3821      EGF-LIKE 14.
FT   DOMAIN    3822   3859      EGF-LIKE 15.
FT   DOMAIN    4146   4182      EGF-LIKE 16.
FT   DOMAIN    4195   4231      EGF-LIKE 17.
FT   DOMAIN    4231   4267      EGF-LIKE 18.
FT   DOMAIN    4267   4303      EGF-LIKE 19.
FT   DOMAIN    4303   4339      EGF-LIKE 20.
FT   DOMAIN    4339   4374      EGF-LIKE 21.
FT   DOMAIN    4372   4409      EGF-LIKE 22.
FT   SITE      3939   3942      RECOGNITION SITE FOR PROTEOLYTICAL
FT                              PROCESSING (POTENTIAL).
FT   SITE      4472   4472      CRITICAL FOR ENDOCYTOSIS (BY SIMILARITY).
FT   SITE      4506   4506      CRITICAL FOR ENDOCYTOSIS (BY SIMILARITY).
FT   DISULFID    29     42      BY SIMILARITY.
FT   DISULFID    36     55      BY SIMILARITY.
FT   DISULFID    49     66      BY SIMILARITY.
FT   DISULFID    74     87      BY SIMILARITY.
FT   DISULFID    81    100      BY SIMILARITY.
FT   DISULFID    94    110      BY SIMILARITY.
FT   DISULFID   117    126      BY SIMILARITY.
FT   DISULFID   122    135      BY SIMILARITY.
FT   DISULFID   137    150      BY SIMILARITY.
FT   DISULFID   156    166      BY SIMILARITY.
FT   DISULFID   162    175      BY SIMILARITY.
FT   DISULFID   177    190      BY SIMILARITY.
FT   DISULFID   480    495      BY SIMILARITY.
FT   DISULFID   491    506      BY SIMILARITY.
FT   DISULFID   508    521      BY SIMILARITY.
FT   DISULFID   805    816      BY SIMILARITY.
FT   DISULFID   812    825      BY SIMILARITY.
FT   DISULFID   827    840      BY SIMILARITY.
FT   DISULFID   852    864      BY SIMILARITY.
FT   DISULFID   859    877      BY SIMILARITY.
FT   DISULFID   871    888      BY SIMILARITY.
FT   DISULFID   893    905      BY SIMILARITY.
FT   DISULFID   900    918      BY SIMILARITY.
FT   DISULFID   912    929      BY SIMILARITY.
FT   DISULFID   934    946      BY SIMILARITY.
FT   DISULFID   941    959      BY SIMILARITY.
FT   DISULFID   953    969      BY SIMILARITY.
FT   DISULFID   974    987      BY SIMILARITY.
FT   DISULFID   982   1000      BY SIMILARITY.
FT   DISULFID   994   1009      BY SIMILARITY.
FT   DISULFID  1013   1025      BY SIMILARITY.
FT   DISULFID  1020   1038      BY SIMILARITY.
FT   DISULFID  1032   1049      BY SIMILARITY.
FT   DISULFID  1060   1073      BY SIMILARITY.
FT   DISULFID  1067   1086      BY SIMILARITY.
FT   DISULFID  1080   1095      BY SIMILARITY.
FT   DISULFID  1102   1116      BY SIMILARITY.
FT   DISULFID  1110   1129      BY SIMILARITY.
FT   DISULFID  1123   1138      BY SIMILARITY.
FT   DISULFID  1143   1157      BY SIMILARITY.
FT   DISULFID  1150   1170      BY SIMILARITY.
FT   DISULFID  1164   1180      BY SIMILARITY.
FT   DISULFID  1183   1194      BY SIMILARITY.
FT   DISULFID  1190   1204      BY SIMILARITY.
FT   DISULFID  1206   1219      BY SIMILARITY.
FT   DISULFID  1225   1235      BY SIMILARITY.
FT   DISULFID  1231   1244      BY SIMILARITY.
FT   DISULFID  1246   1259      BY SIMILARITY.
FT   DISULFID  1538   1551      BY SIMILARITY.
FT   DISULFID  1547   1561      BY SIMILARITY.
FT   DISULFID  1563   1576      BY SIMILARITY.
FT   DISULFID  1846   1857      BY SIMILARITY.
FT   DISULFID  1853   1867      BY SIMILARITY.
FT   DISULFID  1869   1882      BY SIMILARITY.
FT   DISULFID  2155   2166      BY SIMILARITY.
FT   DISULFID  2162   2176      BY SIMILARITY.
FT   DISULFID  2178   2190      BY SIMILARITY.
FT   DISULFID  2476   2487      BY SIMILARITY.
FT   DISULFID  2483   2497      BY SIMILARITY.
FT   DISULFID  2499   2511      BY SIMILARITY.
FT   DISULFID  2518   2531      BY SIMILARITY.
FT   DISULFID  2526   2544      BY SIMILARITY.
FT   DISULFID  2538   2555      BY SIMILARITY.
FT   DISULFID  2560   2572      BY SIMILARITY.
```

```
FT   DISULFID   2567   2585       BY SIMILARITY.
FT   DISULFID   2579   2594       BY SIMILARITY.
FT   DISULFID   2599   2611       BY SIMILARITY.
FT   DISULFID   2606   2624       BY SIMILARITY.
FT   DISULFID   2618   2633       BY SIMILARITY.
FT   DISULFID   2638   2660       BY SIMILARITY.
FT   DISULFID   2654   2673       BY SIMILARITY.
FT   DISULFID   2667   2682       BY SIMILARITY.
FT   DISULFID   2690   2702       BY SIMILARITY.
FT   DISULFID   2697   2715       BY SIMILARITY.
FT   DISULFID   2709   2724       BY SIMILARITY.
FT   DISULFID   2732   2744       BY SIMILARITY.
FT   DISULFID   2739   2757       BY SIMILARITY.
FT   DISULFID   2751   2767       BY SIMILARITY.
FT   DISULFID   2772   2785       BY SIMILARITY.
FT   DISULFID   2779   2798       BY SIMILARITY.
FT   DISULFID   2792   2810       BY SIMILARITY.

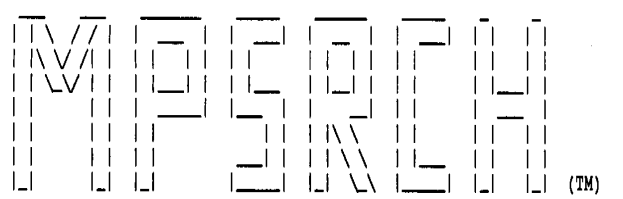
Note: remainder of annotations omitted.

  Query Match          12.9%;  Score 105;  DB 1;  Length 4543;
  Best Local Similarity  34.3%;  Pred. No. 1.16e-05;
  Matches   23;  Conservative   10;  Mismatches  29;  Indels   5;  Gaps   5;

Db   4317 CQMSRDGVKQCRCPPQFEGAQCQ-DNKCSRCQEGKCNINRQSGDVSCICPDGKIAP-SCL 4374
         |::| :|   | | | | | :|| :| |   |  :  | || |  ||    :|::   |
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQENPCLG-QVVREVIRRQKGYASCA-TASKVPIMECR 59

Db   4375 T-CDSYC 4380
          | : |
Qy     60 GGCGPQC 66


Search completed: Fri May 28 09:00:04 1999
Job time : 16 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Psrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:00:22 1999; MasPar time 9.89 Seconds
568.587 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-6
Description: (1-103) from US09191647.pep
Perfect Score: 813
Sequence: 1 QCHISDQGEPYCLCQPGFSG.....GSSFVEEVERHLECGCLACS 103

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 36.755; Variance 55.164; scale 0.666

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	813	100.0	739	4	O75094	MEGF5 (FRAGMENT).	1.40e-189
2	740	91.0	1523	11	O88280	MEGF5.	2.05e-169
3	413	50.8	1531	11	O88279	MEGF4.	1.01e-80
4	307	37.8	79	4	O75093	MEGF4 (FRAGMENT).	4.61e-53
5	176	21.6	601	5	Q20204	F40E10.4 PROTEIN (FRAG	8.68e-21
6	121	14.9	367	11	O54775	ELM1.	1.21e-08
7	120	14.8	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.94e-08
8	116	14.3	381	4	O43775	CYR61 PROTEIN.	1.28e-07
9	116	14.3	955	4	Q99466	NOTCH4 (FRAGMENT).	1.28e-07
10	116	14.3	1999	4	Q99940	NOTCH4.	1.28e-07
11	116	14.3	2003	4	O00306	NOTCH4.	1.28e-07
12	112	13.8	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	8.16e-07
13	112	13.8	2470	11	O35516	CELL SURFACE PROTEIN.	8.16e-07
14	111	13.7	2352	5	O61240	HRNOTCH PROTEIN.	1.29e-06
15	110	13.5	1964	11	O35442	NOTCH4.	2.04e-06
16	110	13.5	2447	13	O13149	NOTCH 2 (FRAGMENT).	2.04e-06
17	109	13.4	434	11	O55139	JAGGED2 PROTEIN (FRAGM	3.22e-06
18	109	13.4	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	3.22e-06
19	109	13.4	1202	11	P97607	JAGGED2 (FRAGMENT).	3.22e-06
20	108	13.3	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	5.07e-06
21	107	13.2	530	5	Q24526	SLIT LOCUS ENCODING A	7.97e-06
22	103	12.7	308	6	O46370	PREADIPOCYTE FACTOR-1.	4.77e-05
23	103	12.7	802	13	O57462	DELTAA.	4.77e-05
24	103	12.7	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	4.77e-05
25	103	12.7	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	4.77e-05
26	102	12.5	387	11	Q06007	NOTCH PROTEIN HOMOLOG	7.42e-05
27	102	12.5	406	5	Q25059	FIBROPELLIN III (FRAGM	7.42e-05
28	101	12.4	717	13	P87357	DELTAD TRANSMEMBRANE P	1.15e-04
29	101	12.4	1042	4	Q13792	APOMUCIN (FRAGMENT).	1.15e-04
30	101	12.4	1081	4	O76065	MUC5AC PROTEIN (FRAGME	1.15e-04
31	101	12.4	2180	5	O01768	SIMILARITY TO EGF-LIKE	1.15e-04
32	101	12.4	2531	5	O16004	NOTCH HOMOLOG.	1.15e-04
33	100	12.3	752	13	O42374	NOTCH RECEPTOR PROTEIN	1.78e-04
34	99	12.2	263	4	Q99740	SOLUBLE PROTEIN JAGGED	2.76e-04
35	99	12.2	529	5	Q25058	FIBROPELLIN IA (FRAGME	2.76e-04
36	99	12.2	728	13	Q90656	TRANSMEMBRANE PROTEIN	2.76e-04
37	99	12.2	1218	4	O14902	TRANSMEMBRANE PROTEIN	2.76e-04
38	99	12.2	1218	4	O15122	JAGGED1.	2.76e-04
39	99	12.2	1218	4	Q15816	TRANSMEMBRANE PROTEIN	2.76e-04
40	99	12.2	1219	11	Q63722	JAGGED PROTEIN.	2.76e-04
41	99	12.2	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	2.76e-04
42	98	12.1	597	11	O35727	FACTOR XII.	4.25e-04
43	98	12.1	762	13	O42373	NOTCH RECEPTOR PROTEIN	4.25e-04
44	98	12.1	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	4.25e-04
45	98	12.1	1372	5	P91526	SIMILARITY TO MULTIPLE	4.25e-04

ALIGNMENTS

RESULT 1
ID O75094 PRELIMINARY; PRT; 739 AA.
AC O75094;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5 (FRAGMENT).
GN MEGF5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011538; D1033429; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 739 AA; 80364 MW; DC6BCB63 CRC32;

Query Match 100.0%; Score 813; DB 4; Length 739;
Best Local Similarity 100.0%; Pred. No. 1.40e-189;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 637 QCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRG 696
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRG 60

Db 697 GCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 739
|||||||||||||||||||||||||||||||||||||||||||||
Qy 61 GCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 103

RESULT 2
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5.
GN   MEGF5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011531; D1033424; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1523 AA;  167767 MW;  2BD845D0 CRC32;

  Query Match           91.0%;  Score 740;  DB 11;  Length 1523;
  Best Local Similarity  88.3%;  Pred. No. 2.05e-169;
  Matches   91;  Conservative    6;  Mismatches   6;  Indels   0;  Gaps   0;

Db   1421 QCHISDRGEPYCLCQPGFSGNHCEQENPCLGEIVREAIRRQKDYASCATASKVPIMVCRG 1480
           ||||||:|||||||||||||:||:|||||||::||| ||||| |||||||||||| |||
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRG 60

Db   1481 GCGSQCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRECS 1523
           |||:||||| ||||||||||||||||||||||||||||||  ||
Qy     61 GCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 103


RESULT    3
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match           50.8%;  Score 413;  DB 11;  Length 1531;
  Best Local Similarity  50.0%;  Pred. No. 1.01e-80;
  Matches   52;  Conservative   20;  Mismatches  31;  Indels   1;  Gaps   1;

Db   1428 HCQASATRGAHCVCSPGFSGELCEQESECRGDPVRDFHRVQRGYAICQTTRPLSWVECRG 1487
           :|: |    : |:| |||||| |:||: | |: ||:  | |:||| | |:  :: :||||
Qy      1 QCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRG 60

Db   1488 ACPGQGCCQGLRLKRRKLTFECSDGTSFAEEVEKPTKCGCAPCA 1531
           :| |  |||  | ||||  |:|:||:|| ||||:   ||| :|:
Qy     61 GC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 103


RESULT    4
ID   O75093      PRELIMINARY;      PRT;  79 AA.
AC   O75093;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4 (FRAGMENT).
GN   MEGF4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011537; D1033428; -.
DR   PROSITE; PS01185; CTCK_1; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   79 AA;  8809 MW;  96C95FFE CRC32;

  Query Match           37.8%;  Score 307;  DB 4;  Length 79;
  Best Local Similarity  49.4%;  Pred. No. 4.61e-53;
  Matches   39;  Conservative   16;  Mismatches  23;  Indels   1;  Gaps   1;

Db      1 ESECRGDPVRDFHQVQRGYAICQTTRPLSWVECRGSCPGQGCCQGLRLKRRKFTFECSDG 60
            |: | |: ||:  : |:||| | |:  :: :||||:| |  |||  | ||||: |:|:||
Qy     26 ENPCLGQVVREVIRRQKGYASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDG 84

Db     61 TSFAEEVEKPTKCGCALCA 79
            :|| ||||:   |||  |:
Qy     85 SSFVEEVERHLECGCLACS 103


RESULT    5
ID   Q20204      PRELIMINARY;      PRT;  601 AA.
AC   Q20204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F40E10.4 PROTEIN (FRAGMENT).
GN   F40E10.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMYE R.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z69792; E1346469; -.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match          21.6%;  Score 176;  DB 5;  Length 601;
  Best Local Similarity  32.6%;  Pred. No. 8.68e-21;
  Matches   30;  Conservative   20;  Mismatches  36;  Indels   6;  Gaps   6;

Db    503 YMCQCDSHFSGEHCD-EKR-I-KCDKQKFRRHHIENECRSVDRIKIAECNGYCGGEQNCC 559
          | | |:: |||||: |:  :    :: :||:     | :  :: | || | ||  | ||
Qy     11 Y-CLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRGGCGP-Q-CC 67

Db    560 TAVKKKQRKVKMICKNGTTKISTVHIIRQCQC 591
            : : |:||    | :|:: :  |    :| |
Qy     68 QPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    6
ID   O54775      PRELIMINARY;      PRT;   367 AA.
AC   O54775;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ELM1.
GN   ELM1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HEN;
RX   MEDLINE; 98119879.
RA   HASHIMOTO Y., SHINDO-OKADA N., TANI M., NAGAMACHI Y., TAKEUCHI K.,
RA   SHIROISHI T., TOMA H., YOKOTA J.;
RT   "Expression of the Elm1 gene, a novel gene of the CCN (connective
RT   tissue growth factor, Cyr61/Cef10, and neuroblastoma overexpressed
RT   gene) family, suppresses In vivo tumor growth and metastasis of
RT   K-1735 murine melanoma cells.";
RL   J. EXP. MED. 187:289-296(1998).
DR   EMBL; AB004873; D1025874; -.
DR   PROSITE; PS01185; CTCK_1; 1.
SQ   SEQUENCE   367 AA;  40702 MW;  1AB35AB9 CRC32;

  Query Match          14.9%;  Score 121;  DB 11;  Length 367;
  Best Local Similarity  33.3%;  Pred. No. 1.21e-08;
  Matches   20;  Conservative   12;  Mismatches  25;  Indels   3;  Gaps   3;

Db    288 AGCVSTRTYRPKYC-GVCTDNRCCIPYKSKTISVDFQCPEGPGFSRQVLWINACFCNLSC 346
          |:| ::     : | | |   :|| | :||     ||| :|::|  :| :   | | |:|
Qy     45 ASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC-LAC 102


RESULT    7
ID   Q19350      PRELIMINARY;      PRT;  1722 AA.
AC   Q19350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO EGF-LIKE REPEATS. NCBI GI: 1125776.
GN   F11C7.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A., VETTER J.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U42839; G1125776; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PFAM; PF00008; EGF; 24.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1722 AA;  188383 MW;  CCFB86B8 CRC32;

  Query Match          14.8%;  Score 120;  DB 5;  Length 1722;
  Best Local Similarity  39.3%;  Pred. No. 1.94e-08;
  Matches   11;  Conservative    7;  Mismatches  10;  Indels   0;  Gaps   0;

Db    380 CKLDAEGEPFCVCEEGFDGPFCEPKSGC 407
          | :  :|||:|:|: || |  |:  : |
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQENPC 29


RESULT    8
ID   O43775      PRELIMINARY;      PRT;   381 AA.
AC   O43775;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN.
GN   CYR61.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MARTINERIE C., VIEGAS-PEQUIGNOT E., NGUYEN V.C., PERBAL B.;
RL   J. CLIN. PATHOL. MOL. PATHOL. 50:130-136(1997).
DR   EMBL; Y11307; E304665; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   381 AA;  42025 MW;  1B18FF1A CRC32;

  Query Match          14.3%;  Score 116;  DB 4;  Length 381;
  Best Local Similarity  33.3%;  Pred. No. 1.28e-07;
  Matches   19;  Conservative   10;  Mismatches  26;  Indels   2;  Gaps   2;

Db    300 YAGCLSVKKYRPKYC-GSCVDGRCCTPQLTRTVKMRFRCEDGETFSKNVMMIQSCKC 355
          ||:| :  |   : | |:|   :|| |  ::  |  |:| || :|  :|     | |
Qy     44 YASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGC 99


RESULT    9
ID   Q99466      PRELIMINARY;      PRT;   955 AA.
AC   Q99466;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4 (FRAGMENT).
GN   NOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97311416.
RA   SUGAYA K., SASANUMA S., NOHATA J., KIMURA T., FUKAGAWA T.,
RA   NAKAMURA Y., ANDO A., INOKO H., IKEMURA T., MITA K.;
```

```
RT   "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
RT   human counterpart gene of mouse proto-oncogene Int3.";
RL   GENE 189:235-244(1997).
DR   EMBL; D86566; D1013803; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 21.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     955    955
SQ   SEQUENCE   955 AA;  100017 MW;  28507B36 CRC32;

  Query Match          14.3%;  Score 116;  DB 4;  Length 955;
  Best Local Similarity  46.9%;  Pred. No. 1.28e-07;
  Matches   15;  Conservative    5;  Mismatches  10;  Indels   2;  Gaps   2;

Db    406 QCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 437
          ||  :   |  ||||||:||  |:|: : ||
Qy      1 QCHISD-QGEPYCLCQPGFSGEHCQQE-NPCL 30


ULT   10
   Q99940      PRELIMINARY;      PRT;  1999 AA.
AC   Q99940;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   NOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LI L., HUANG G., BANTA A., DENG Y., CHEN L., PHAM Q., ROWEN L.,
RA   HOOD L.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U89335; G1841543; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 5.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1999 AA;  209134 MW;  0680278E CRC32;

  Query Match          14.3%;  Score 116;  DB 4;  Length 1999;
  Best Local Similarity  46.9%;  Pred. No. 1.28e-07;
atches   15;  Conservative    5;  Mismatches  10;  Indels   2;  Gaps   2;

Db    405 QCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 436
          ||  :   |  ||||||:||  |:|: : ||
Qy      1 QCHISD-QGEPYCLCQPGFSGEHCQQE-NPCL 30


RESULT   11
ID   O00306      PRELIMINARY;      PRT;  2003 AA.
AC   O00306;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   HNOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW, HEART;
RA   LI L., HUANG G., BANTA A., YU D., ROWEN L., HOOD L.;
RL   SUBMITTED (MAR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U95299; G2072309; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 5.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2003 AA;  209620 MW;  518CFE96 CRC32;

  Query Match          14.3%;  Score 116;  DB 4;  Length 2003;
  Best Local Similarity  46.9%;  Pred. No. 1.28e-07;
  Matches   15;  Conservative    5;  Mismatches  10;  Indels   2;  Gaps   2;

Db    406 QCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 437
          ||  :   |  ||||||:||  |:|: : ||
Qy      1 QCHISD-QGEPYCLCQPGFSGEHCQQE-NPCL 30


RESULT   12
ID   Q61204      PRELIMINARY;      PRT;  1687 AA.
AC   Q61204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH2-LIKE (EGF REPEAT TRANSMEMBRANE PROTEIN).
GN   NOTCH2L.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=WHOLE EMBRYO;
RA   SELL C., HOFF III H.B.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U57368; G1336628; -.
DR   MGD; MGI:1202397; NOTCH2L.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PFAM; PF00008; EGF; 7.
KW   TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1687 AA;  188528 MW;  73B9DDDC CRC32;

  Query Match          13.8%;  Score 112;  DB 11;  Length 1687;
  Best Local Similarity  40.6%;  Pred. No. 8.16e-07;
  Matches   13;  Conservative    9;  Mismatches  8;  Indels   2;  Gaps   2;

Db    350 CHMLSR-DTYECTCQVGFTGKQCQWTDACLSH 380
          ||:  : : | | || ||:| :||  ::||::
Qy      2 CHISDQGEPY-CLCQPGFSGEHCQQENPCLGQ 32


RESULT   13
ID   O35516      PRELIMINARY;      PRT;  2470 AA.
AC   O35516;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CELL SURFACE PROTEIN.
GN   NOTCH2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RA   HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL   SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D32210; D1022953; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match          13.8%;  Score 112;  DB 11;  Length 2470;
  Best Local Similarity  40.6%;  Pred. No. 8.16e-07;
  Matches   13;  Conservative   9;  Mismatches   8;  Indels   2;  Gaps   2;

Db    121 CHMLSR-DTYECTCQVGFTGKQCQWTDACLSH 151
          ||:  : : | | || ||:| :||  ::||::
Qy      2 CHISDQGEPY-CLCQPGFSGEHCQQENPCLGQ 32


RESULT   14
ID   O61240      PRELIMINARY;      PRT;  2352 AA.
AC   O61240;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HRNOTCH PROTEIN.
GN   HRNOTCH.
OS   HALOCYNTHIA RORETZI (SEA SQUIRT).
OC   EUKARYOTA; METAZOA; CHORDATA; UROCHORDATA; ASCIDIACEA; STOLIDOBRANCHIA;
OC   PYURIDAE; HALOCYNTHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HORI S., SAITOH T., MATSUMOTO M., MAKABE K.W., NISHIDA H.;
RL   DEV. GENES EVOL. 207:371-380(1997).
DR   EMBL; AB001327; D1026501; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS01186; EGF_2; 22.
DR   PROSITE; PS01187; EGF_CA; 18.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2352 AA;  252623 MW;  816976D4 CRC32;

  Query Match          13.7%;  Score 111;  DB 5;  Length 2352;
  Best Local Similarity  54.8%;  Pred. No. 1.29e-06;
  Matches   17;  Conservative   5;  Mismatches   5;  Indels   4;  Gaps   4;

Db    846 QC-LDDVG-SYKCLCLPGFEGNNCQEEVNEC 874
          || : | | :| ||| ||| |::||:| | |
Qy      1 QCHISDQGEPY-CLCQPGFSGEHCQQE-NPC 29


RESULT   15
ID   O35442      PRELIMINARY;      PRT;  1964 AA.
AC   O35442;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   NOTCH4.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROWEN L., MAHAIRAS G., QIN S., AHEARN M.E., DANKERS C., LASKY S.,
RA   LORETZ C., SCHMIDT S., TIPTON S., TRAICOFF R., ZACKRONE K., HOOD L.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF030001; G2564947; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1964 AA;  206699 MW;  CE2CA3B6 CRC32;

  Query Match          13.5%;  Score 110;  DB 11;  Length 1964;
  Best Local Similarity  43.8%;  Pred. No. 2.04e-06;
  Matches   14;  Conservative   7;  Mismatches   9;  Indels   2;  Gaps   2;

Db    449 C-INTPGSFNCLCLPGYTGSRCEADHNECLSQ 479
          | |:  |   ||| ||::| :|: : | ||:|
Qy      2 CHISDQGEPYCLCQPGFSGEHCQQE-NPCLGQ 32


Search completed: Fri May 28 09:00:54 1999
Job time : 32 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:02:39 1999; MasPar time 8.30 Seconds
397.227 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-8
Description: (1-155) from US09191647.pep
Perfect Score: 1092
Sequence: 1 RNPXICDCNLQWLAQINLQK.....FINDKSFEKLSKLRELXLND 155

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

atistics: Mean 30.771; Variance 136.870; scale 0.225

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	469	42.9	1480	5	R25079	Drosophila SLIT prote	4.43e-32
2	403	36.9	1534	30	W46966	Amino acid sequence o	5.07e-26
3	197	18.0	369	15	R87951	Rat neurotrophic bigl	1.02e-07
4	193	17.7	139	8	R42263	Decorin sequence PT-7	2.24e-07
5	193	17.7	186	8	R42264	Decorin sequence PT-7	2.24e-07
6	193	17.7	234	8	R42265	Decorin sequence PT-7	2.24e-07
7	193	17.7	280	8	R42266	Decorin sequence PT-7	2.24e-07
8	193	17.7	305	8	R42267	Decorin sequence PT-7	2.24e-07
9	193	17.7	331	8	R42260	Mature decorin PT-65.	2.24e-07
10	193	17.7	332	15	R87953	Bovine neurotrophic b	2.24e-07
11	193	17.7	342	17	R89439	Human recombinant dec	2.24e-07
12	193	17.7	353	1	R05160	Sequence of human bon	2.24e-07
13	193	17.7	369	15	R87952	Human neurotrophic bi	2.24e-07
14	193	17.7	1388	18	R89471	Collagen/decorin fusi	2.24e-07
15	192	17.6	196	5	R29102	Drosophila SLIT prote	2.72e-07
16	180	16.5	368	1	R05159	Sequence of human bon	2.78e-06
17	161	14.7	94	8	R42262	Decorin sequence PT-7	1.05e-04
18	161	14.7	1091	27	W41641	Sequence used in dete	1.05e-04
19	142	13.0	111	30	W58846	Human AS209_1 secrete	3.66e-03
20	140	12.8	433	18	R98454	Oligodendrocyte-myeli	5.30e-03
21	139	12.7	345	23	W09405	Pineal gland specific	6.37e-03
22	139	12.7	560	12	R71294	Human glycoprotein V.	6.37e-03
23	136	12.5	661	39	W87556	B cell surface protei	1.10e-02
24	136	12.5	661	25	W28510	Product of clone J422	1.10e-02
25	136	12.5	661	28	W47274	Human B-cell activati	1.10e-02
26	136	12.5	4303	17	R90302	Polycystic kidney dis	1.10e-02
27	134	12.3	605	17	R85888	WD-40 domain-contg. i	1.59e-02
28	132	12.1	245	28	W39170	Human PKD1 protein fr	2.29e-02
29	132	12.1	455	28	W39171	Human PKD1 protein fr	2.29e-02
30	132	12.1	784	39	W86350	Human DNAX toll-like	2.29e-02
31	132	12.1	784	39	W90069	Human TNF-alpha conve	2.29e-02
32	132	12.1	784	29	W48245	Human pro-tumour necr	2.29e-02
33	132	12.1	4302	19	W00870	Polycystic kidney dis	2.29e-02
34	132	12.1	4302	29	W33396	Human PKD1 polypeptid	2.29e-02
35	132	12.1	4302	28	W23830	Human PKD1 protein.	2.29e-02
36	132	12.1	4339	15	R75916	Polycystic kidney dis	2.29e-02
37	132	12.1	4339	19	R87539	Polycystic kidney dis	2.29e-02
38	131	12.0	390	20	W06532	Gonadotropin receptor	2.74e-02
39	124	11.4	695	5	R27558	FSHR.	9.69e-02
40	124	11.4	1045	39	W86354	Human DNAX toll-like	9.69e-02
41	123	11.3	904	39	W86351	Human DNAX toll-like	1.16e-01
42	122	11.2	634	6	R30520	N-terminal of LH rece	1.38e-01
43	122	11.2	692	6	R30503	N-terminal of LH rece	1.38e-01
44	122	11.2	695	6	R30524	N-terminal of LH rece	1.38e-01
45	122	11.2	695	6	R30506	N-terminal of LH rece	1.38e-01

ALIGNMENTS

RESULT 1
ID R25079 standard; Protein; 1480 AA.
AC R25079;
DT 05-JAN-1993 (first entry)
DE Drosophila SLIT protein involved in axon pathway development.
KW Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW midline glial cells; axonogenesis; cell-cell interaction; ss.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT peptide 1..36
FT /label= signal
FT domain 73..294
FT /label= Flank_LRR_Flank_1
FT /note= "mediates adhesive events"
FT domain 295..518
FT /label= Flank-LRR-Flank_2
FT /note= "mediates adhesive events"
FT domain 519..714
FT /label= Flank_LRR_Flank_3
FT /note= "mediates adhesive events"
FT domain 715..910
FT /label= Flank_LRR_Flank_4
FT /note= "mediates adhesive events"
FT region 911..1150
FT /label= Tandem_EGF_like_repeats
FT /note= "involved in protein-protein interactions"
FT region 1353..1393
FT /label= 7th_EGF_like_repeat
FT /note= "involved in receptor-ligand interactions"
FT region 1394..1404
FT /label= alternative_splice_segment
FT /note= "developmentally regulated"
FT region 1405..1480
FT /label= C-terminal_region
PN WO9210518-A.
PD 25-JUN-1992.
PF 27-NOV-1991; U09055.
PR 07-DEC-1990; US-624135.
PA (UYYA) UNIV YALE.

PI Artavanis-Tsakonas S, Rothberg JM;
DR WPI; 92-234590/28.
DR N-PSDB; Q25811.
PT SLIT protein and sequence elements for treating
PT neuro-degenerative disease - useful for Alzheimer's disease,
PT nerve damage and Parkinson's disease, for diagnosis of cancer
PS Claim 1; Page 84-89; 122pp; English.
CC The SLIT protein is necessary for normal development of the midline
CC of the CNS, partic. the midline glial cells, and for the
CC concomitant formation of the commisural axon pathways. The process
CC is dependent on the level of SLIT protein expression. It appears
CC that SLIT protein is excreted by the midline glial cells where it
CC is synthesised and is eventually associated with the surfaces of
CC axons that traverse tham. The SLIT protein is tightly localised to
CC the muscle attachment sites and to the sites of contact between
CC adjacent pairs of cardioblasts as they coalesce to form the lumen
CC of the larval heart. The SLIT protein defines a new set of
CC molecules (TAGONS) which play a key role in axon outgrowth and
CC pathfinding. SLIT can be used as a nerve regenerative in
CC neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC region-Flank" domains, the C-terminal region and part of the
CC alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC and at least 1 EGF-like repeat element from the SLIT protein.
CC See also R29102.
SQ Sequence 1480 AA;

Query Match 42.9%; Score 469; DB 5; Length 1480;
Best Local Similarity 42.9%; Pred. No. 4.43e-32;
Matches 67; Conservative 42; Mismatches 44; Indels 3; Gaps 3;

```
Db   451 knpficdcnlrwvad-ylhknpietsgarcespkrmhrrriesvreekfkcswgelrmkl 509
           :|| ||||||:|:|:  |:|| |||||||| |||:::::: ::  :|||| :|  : :
Qy     1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSESFVSM 59

Db   510 sge-crmdsdcpamchcegttvdctgrrlkeiprdiplhttelllndnelgrissdglfg 568
           :: | :|| ||: | | |||||  | |: ||  ||  :|:|||: |::: :  :: :
Qy    60 YADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIH 119

Db   569 rlphlvklelkrnqltgiepnafegashiqelqlge 604
           | :|  |:|  |::| |: ::||  | ::|| | :
Qy   120 VLENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155
```

RESULT 2
W46966 standard; Protein; 1534 AA.
W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 36.9%; Score 403; DB 30; Length 1534;
Best Local Similarity 39.4%; Pred. No. 5.07e-26;
Matches 63; Conservative 40; Mismatches 49; Indels 8; Gaps 7;

```
Db   438 qnpficdcnlkwladf-lrtnpietsgarcasprrlankrigqikskkfrcsakeqyfip 496
           :|| |||||| |||:: |: | |||||||| |:|| :|::: : ::||:| : |  |::
Qy     1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSES-FVS 58

Db   497 gtedyqlnsecnsdvvcphkcrceanvvecsslkltkiperipqstaelrlnnneisile 556
                | :| |  | :||  | | :  |:|:   |  ||  ||: :::| |: |:|| ::
Qy    59 -M--YA-DS-CFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVD 113

Db   557 atgmfkklthlkkinlsnnkvseiedgafegaasvselhl 596
              : :  | :|  ::|||| :: |:| :||  : : || |
Qy   114 LNSNIHVLENLEXLDLSNNHITFINDKSFEKLSKLRELXL 153
```

RESULT 3
ID R87951 standard; Protein; 369 AA.
AC R87951;
DT 20-MAR-1996 (first entry)
DE Rat neurotrophic biglycan.
KW Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
KW learning.
OS Rattus sp.
FH Key Location/Qualifiers
FT peptide 1..37
FT /label= Sig_peptide
FT region 44..60
FT /label= Hypervariable_region
PN WO9530432-A1.
PD 16-NOV-1995.
PF 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Hasenoehrl R, Huston J, Junghans U, Kappler J, Koops A;
PI Mueller HW;
DR WPI; 95-403938/51.
DR N-PSDB; T08768.
PT Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s) -
PT for maintain structural and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 1; Page 44-45; 60pp; English.
CC Rat biglycan (R87951) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. Recombinant biglycan,
CC obtd. by expression of encoding cDNA (T08768) in eukaryotic host
CC cells, can be used to enhance the survival and maintain the structure
CC and function of CNS neurons during normal ageing as well as after
CC pathological and/or traumatic nervous system damage. It can also
CC be used to restore function following nervous system lesions and
CC degenerative diseases, and to improve learning efficiency and memory
CC in the elderly and in patients with dementia.
SQ Sequence 369 AA;

Query Match 18.0%; Score 197; DB 15; Length 369;
Best Local Similarity 34.8%; Pred. No. 1.02e-07;
Matches 31; Conservative 23; Mismatches 34; Indels 1; Gaps 1;

```
Db    60 fsamcpfgchchlrvvqcsdlglktvpkeispdttlldlqnndiselr-kddfkglqhly 118
           : ::||  | |    |:|:  ||:|:|  |:  :| | |  |:|| :  : ::  |::|
Qy    66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db   119 alvlvnnkiskihekafsplrklqklyis 147
              | | || |: |::|:|  | ||: | ::
Qy   126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154
```

RESULT 4
ID R42263 standard; Protein; 139 AA.
AC R42263;
DT 28-APR-1994 (first entry)
DE Decorin sequence PT-74 (N-terminal to LRR4).
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50049.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity
PS Claim 10; Page 41; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 139 AA;

Query Match 17.7%; Score 193; DB 8; Length 139;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 5
ID R42264 standard; Protein; 186 AA.
AC R42264;
DT 28-APR-1994 (first entry)
DE Decorin sequence PT-75 (N-terminal to LRR6).
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50050.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity
PS Claim 10; Page 43-44; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 186 AA;

Query Match 17.7%; Score 193; DB 8; Length 186;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 6
ID R42265 standard; Protein; 234 AA.
AC R42265;
DT 28-APR-1994 (first entry)
DE Decorin sequence PT-76 (N-terminal to LRR8).
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50051.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity
PS Claim 10; Page 45-46; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 234 AA;

Query Match 17.7%; Score 193; DB 8; Length 234;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 7
ID R42266 standard; Protein; 280 AA.
AC R42266;
DT 28-APR-1994 (first entry)
DE Decorin sequence PT-77 (N-terminal to LRR10).
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50052.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity

PS Claim 10; Page 47-48; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 280 AA;

Query Match 17.7%; Score 193; DB 8; Length 280;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 8
ID R42267 standard; Protein; 305 AA.
AC R42267;
DT 28-APR-1994 (first entry)
DE Decorin sequence PT-78 (N-terminal to half C-terminal).
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50053.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity
PS Claim 10; Page 49-50; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 305 AA;

Query Match 17.7%; Score 193; DB 8; Length 305;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 9
ID R42260 standard; Protein; 331 AA.
AC R42260;
DT 28-APR-1994 (first entry)
DE Mature decorin PT-65.
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW fusion protein; maltose binding protein; tumour growth; inhibition;
KW decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
FH Key Location/Qualifiers
FT region 1..45
FT /label= N-terminal_region
FT /note= "contains 4 Cys residues"
FT region 46..280
FT /label= repeat_region
FT /note= "contains 10 leucine-rich repeats"
FT region 281..331
FT /label= C-terminal_region
PN WO9320202-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03171.
PR 03-APR-1992; US-865652.
PA (LJOL-) LA JOLLA CANCER RES FOUND.
PI Cardenas J, Craig W, Mullen DG, Pierschbacher MD;
PI Ruoslahti EI;
DR WPI; 93-336910/42.
DR N-PSDB; Q50046.
PT Active fragments of protein esp. decorin - with cell regulatory
PT factor domain, useful for inhibiting cell regulatory factor
PT activity
PS Claim 10; Page 36-38; 77pp; English.
CC Active fragments of decorin (full-length coding sequence Q50046)
CC were generated by PCR and fused to Maltose Binding Protein. The
CC resulting fusion proteins were useful for inhibiting the activity of
CC a cell regulatory factor, esp. TGF-beta, and hence for treating
CC conditions associated with over-activity of the growth factor such
CC as certain tumours.
SQ Sequence 331 AA;

Query Match 17.7%; Score 193; DB 8; Length 331;
Best Local Similarity 32.6%; Pred. No. 2.24e-07;
Matches 29; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 22 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 80
: ::|| :|:| |:|: ||: :| :| :| | | |:|: : :::: | ||
Qy 66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db 81 alilvnnkiskvspgaftplvklerlyls 109
| | || |: :: :| | || | |:
Qy 126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

RESULT 10
ID R87953 standard; Protein; 332 AA.
AC R87953;
DT 20-MAR-1996 (first entry)
DE Bovine neurotrophic biglycan.
KW Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
KW learning.
OS Bos taurus.
FH Key Location/Qualifiers
FT region 7..23
FT /label= Hypervariable_region
PN WO9530432-A1.
PD 16-NOV-1995.
PF 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Hasenoehrl R, Huston J, Junghans U, Kappler J, Koops A;
PI Mueller HW;
DR WPI; 95-403938/51.
PT Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s) -
PT for maintain structural and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 3; Fig 8; 60pp; English.
CC Bovine biglycan (R87953) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. It can be used to enhance
CC the survival and maintain the structure and function of CNS neurons
CC during normal ageing as well as after pathological and/or traumatic
CC nervous system damage. It can also be used to restore function
CC following nervous system lesions and degenerative diseases, and to
CC improve learning efficiency and memory in the elderly and in patients

```
CC  with dementia.
SQ  Sequence  332 AA;

  Query Match          17.7%;  Score 193;  DB 15;  Length 332;
  Best Local Similarity  34.5%;  Pred. No. 2.24e-07;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     25 amcpfgchchlrvvqcsdlglkavpkeispdttlldlqnndiselr-kddfkglqhlyal 83
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db     84 vlvnnkiskihekafsplrklqklyis 110
            | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  11
    R89439 standard; Protein; 342 AA.
    R89439;
DT  20-AUG-1996  (first entry)
DE  Human recombinant decorin.
KW  Decorin; PG-II; PG-40; proteoglycan; guanidinium ion.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  peptide         1..14
FT                  /label= Sig_peptide
PN  WO9601842-A1.
PD  25-JAN-1996.
PF  07-JUL-1995; U08542.
PR  08-JUL-1994; US-272919.
PA  (LJOL-) LA JOLLA CANCER RES FOUND.
PI  Craig WS,  Harper JR,  Hernandez SD,  Kostel PJ,  Parker JR;
PI  Vedvick TS;
DR  WPI; 96-097586/10.
DR  N-PSDB; T10741.
PT  Purificn. of human recombinant decorin - using a strong anion
PT  exchange resin, a hydrophobic interaction chromatography resin and a
PT  strong anion exchange resin
PS  Disclosure; Fig 1A-D; 55pp; English.
CC  Human recombinant decorin (R89439) was obtd. by expression of a
CC  cDNA clone (T10741) in CHO host cells.  Decorin (or PGII or PG-40)
CC  is a proteoglycan having a 40 kDa core protein.  Recombinant
CC  decorin can be produced by cotransfection of CHO-DG44 cells with
CC  pSV2-decorin and pSV2dhfr.  Large-scale cultures can be performed
CC  using CHO cells attached to microcarrier beads.  The recombinant
    protein is purified from the cells using a 3-step chromatographic
    procedure.  It can be used for the highly sensitive detection of
    guanidinium ions (ppm range), partic. in protein-contg. solns.
CC  purified using GuHCl, and also has therapeutic applns.
SQ  Sequence  342 AA;

  Query Match          17.7%;  Score 193;  DB 17;  Length 342;
  Best Local Similarity  32.6%;  Pred. No. 2.24e-07;
  Matches   29;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db     34 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 92
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::: | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db     93 alilvnnkiskvspgaftplvklerlyls 121
            | | || |: ::  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  12
ID  R05160 standard; protein; 353 AA.
AC  R05160;
DT  09-OCT-1990  (first entry)
DE  Sequence of human bone proteoglycan II (decorin).
KW  Osteoporosis; rheumatoid arthritis; Paget's disease;
KW  atherosclerosis; periodontal; human bone matrix; proteoglycan.
OS  Homo sapiens.
PN  US7432044-A.
PD  17-APR-1990.
PF  3-NOV-1989; 432044.
PR  3-NOV-1989; US-432044.
PA  (USSH) Nat Inst of Health.
PI  Termine J;
DR  WPI; 90-178641/23.
DR  N-PSDB; Q04491.
PT  Human bone matrix DNA and proteins -
PT  used in detection, diagnosis and treatment involving skeletal
PT  and/or connective tissue disease states.
PS  Disclosure; p; English.
CC  Probes and Abs raised to the proteins can be used to determine
CC  their levels useful in diagnosis of associated conective tissue
CC  diseases states such as osteoporosis, osteo/rheumatoid arthritis,
CC  Paget's disease, artherosclerosis and periodontal disease.
CC  Proteins may also be used to induce or block biological function.
SQ  Sequence  353 AA;

  Query Match          17.7%;  Score 193;  DB 1;  Length 353;
  Best Local Similarity  32.6%;  Pred. No. 2.24e-07;
  Matches   29;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db     44 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 102
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::: | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    103 alilvnnkiskvspgaftplvklerlyls 131
            | | || |: ::  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  13
ID  R87952 standard; Protein; 369 AA.
AC  R87952;
DT  20-MAR-1996  (first entry)
DE  Human neurotrophic biglycan.
KW  Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW  neurotrophic; central nervous system; CNS; memory loss; dementia;
KW  learning.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  peptide         1..37
FT                  /label= Sig_peptide
FT  region          44..60
FT                  /label= Hypervariable_region
PN  WO9530432-A1.
PD  16-NOV-1995.
PF  09-MAY-1994; E01479.
PR  09-MAY-1994; WO-E01479.
PA  (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI  Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI  Mueller HW;
DR  WPI; 95-403938/51.
PT  Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s) -
PT  for maintain structural and function of the CNS and attenuating
PT  memory deficit(s) in the elderly and patients with dementia
PS  Claim 3; Fig 8; 60pp; English.
CC  Human biglycan (R87952) is a chondroitin sulphate proteoglycan with
CC  neurotrophic activity for brain neurons.  It can be used to enhance
CC  the survival and maintain the structure and function of CNS neurons
CC  during normal ageing as well as after pathological and/or traumatic
CC  nervous system damage.  It can also be used to restore function
CC  following nervous system lesions and degenerative diseases, and to
CC  improve learning efficiency and memory in the elderly and in patients
CC  with dementia.
SQ  Sequence  369 AA;

  Query Match          17.7%;  Score 193;  DB 15;  Length 369;
  Best Local Similarity  34.5%;  Pred. No. 2.24e-07;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     62 amcpfgchchlrvvqcsdlglksvpkeispdttlldlqnndiselr-kddfkglqhlyal 120
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
```

```
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    121 vlvnnkiskihekafsplrklqklyis 147
          | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   14
ID   R89471 standard; Protein; 1388 AA.
AC   R89471;
DT   01-OCT-1996  (first entry)
DE   Collagen/decorin fusion protein.
KW   Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW   bone formation; tissue repair; fusion protein.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   domain          1..1057
FT                   /label= Collagen-IA
FT                   /note= "collagen IA alpha-helical domain"
FT   peptide         1058..1059
FT                   /label= Linker_peptide
FT   domain          1060..1388
FT                   /label= Decorin
FT   misc_difference 887
FT                   /note= "unidentified amino acid"
FT   misc_difference 890
FT                   /note= "unidentified amino acid"
PN   CA2151547-A.
PD   11-DEC-1995.
PF   12-JUN-1995; 151547.
PR   10-JUN-1994; US-259263.
PA   (USSU ) US SURGICAL CORP.
PI   Espino P,  Gruskin EA;
DR   WPI; 96-140144/15.
DR   N-PSDB; T16517.
PT   Chimaeric DNA encoding protein contg. extracellular matrix protein
PT   domain  - and cellular regulatory factor domain, partic. useful as
PT   osteogenic agents, also related vectors, transformed cells and
PT   chimaeric proteins.
PS   Disclosure; Fig 7; 59pp; English.
CC   A fusion protein (R89471) comprises the alpha-helical region of
CC   human collagen I(a) linked to human dermatan sulphate proteoglycan
CC   (decorin).  It can be expressed in Escherichia coli transformants
CC   carrying a vector incorporating a chimeric gene (T16517) coding for
CC   the fusion.  The decorin binds to type I collagen and thus affects
CC   Elbril formation.  It inhibits the cell attachment-promoting
CC   activity of collagen and fibrinogen by binding to such molecules
CC   near their cell binding sites.  The collagen moiety provides an
CC   integral substratum or scaffolding for the decorin.  The fusion
CC   protein acts to reduce scarring of healing tissue.
SQ   Sequence   1388 AA;

  Query Match          17.7%;  Score 193;  DB 18;  Length 1388;
  Best Local Similarity  32.6%;  Pred. No. 2.24e-07;
  Matches   29;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db   1079 lgpvcpfrcqchlrvvqcsdlgldkvpkdlppdttlldlqnnkiteik-dgdfknlknlh 1137
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  ::::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db   1138 alilvnnkiskvspgaftplvklerlyls 1166
          | | || |: ::  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154

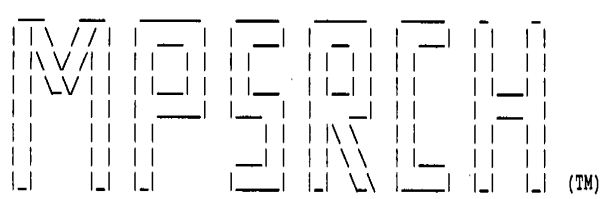

RESULT   15
ID   R29102 standard; Protein; 196 AA.
AC   R29102;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein Flank-LRR-Flank consensus sequence.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; midline glial cells;
KW   axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   region          1..32
FT                   /label= amino_flanking_region
FT   region          33..135
FT                   /label= Leucine_rich_repeat_region
FT   region          136..196
FT                   /label= carboxy_flanking_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 5; Page 95; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT contains 4 Flank-LRR-Flank domains (see R25079)
CC   which mediate adhesive events and this consensus sequence is based
CC   on them. Each of the 4 individual domain sequences is individually
CC   claimed. See also Q25811.
SQ   Sequence   196 AA;

  Query Match          17.6%;  Score 192;  DB 5;  Length 196;
  Best Local Similarity  31.0%;  Pred. No. 2.72e-07;
  Matches   26;  Conservative   4;  Mismatches  53;  Indels   1;  Gaps   1;

Db     1 cpxxcxcxgxxvdcxxxqlxxxpxxxpxdttxxxxxxnxixxlxxxx-fxxlxxlxxlxl 59
           ||  | | |  |||    |  |  |  :|     | |  :    :  |  | || |
Qy    70 CPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXLDL 129

Db    60 xxnxixxlxxxxfxxlxxlxxlil 83
            | |  :   |  |  |  | |
Qy   130 SNNHITFINDKSFEKLSKLRELXL 153


Search completed: Fri May 28 09:03:40 1999
Job time : 61 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:03:57 1999; MasPar time 9.06 Seconds
685.804 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-8
Description: (1-155) from US09191647.pep
Perfect Score: 1092
Sequence: 1 RNPXICDCNLQWLAQINLQK.....FINDKSFEKLSKLRELXLND 155

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 43.006; Variance 90.646; scale 0.474

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	483	44.2	1469	2	B36665	slit protein 2 precur	6.23e-67
2	483	44.2	1480	2	A36665	slit protein 1 precur	6.23e-67
3	201	18.4	420	2	A53531	oncofetal trophoblast	3.72e-16
4	197	18.0	369	2	S20811	proteoglycan I - mous	1.69e-15
5	197	18.0	369	2	S32793	biglycan precursor -	1.69e-15
6	195	17.9	360	2	S06280	decorin precursor - b	3.59e-15
7	195	17.9	360	2	I47020	decorin - rabbit	3.59e-15
8	193	17.7	359	1	NBHUC8	decorin precursor - h	7.61e-15
9	193	17.7	368	1	BGHUN	biglycan precursor -	7.61e-15
10	193	17.7	369	2	S32559	biglycan precursor -	7.61e-15
11	192	17.6	357	2	S24317	decorin precursor - c	1.11e-14
12	184	16.8	322	2	S72271	proteoglycan Lb precu	2.18e-13
13	184	16.8	354	2	A55454	decorin precursor - m	2.18e-13
14	183	16.8	354	2	S29145	decorin precursor - r	3.16e-13
15	182	16.7	316	2	A41781	proteoglycan-Lb - chi	4.57e-13
16	162	14.8	208	2	JN0638	platelet glycoprotein	6.54e-10
17	161	14.7	1091	2	A58532	glial cell membrane g	9.34e-10
18	159	14.6	206	1	NBHUIB	platelet glycoprotein	1.90e-09
19	159	14.6	343	2	A41748	lumican precursor - c	1.90e-09
20	159	14.6	411	1	I55604	platelet glycoprotein	1.90e-09
21	159	14.6	907	2	JE0176	orphan G protein-coup	1.90e-09
22	156	14.3	1535	2	S46224	peroxidasin - fruit f	5.48e-09
23	147	13.5	662	2	S42799	garp precursor - huma	1.26e-07
24	144	13.2	661	2	I56258	RP105 - mouse	3.55e-07
25	142	13.0	440	2	A47530	oligodendrocyte-myeli	7.02e-07
26	140	12.8	440	2	A39613	oligodendrocyte-myeli	1.38e-06
27	139	12.7	230	2	I46918	leucine-rich glycopro	1.94e-06
28	139	12.7	560	2	A60164	platelet membrane gly	1.94e-06
29	136	12.5	682	2	A43318	connectin precursor -	5.32e-06
30	136	12.5	682	2	A49121	cell-surface molecule	5.32e-06
31	134	12.3	605	2	A41915	insulin-like growth f	1.04e-05
32	132	12.1	298	2	JC4130	osteoglycin precursor	2.01e-05
33	132	12.1	4302	2	A38971	polycystic kidney dis	2.01e-05
34	131	12.0	1134	1	A29944	chaoptin precursor -	2.80e-05
35	130	11.9	298	2	B35272	osteoinductive factor	3.89e-05
36	130	11.9	299	2	A35272	osteoinductive factor	3.89e-05
37	130	11.9	338	2	S52284	lumicon, secretory in	3.89e-05
38	130	11.9	380	2	S71876	fibromodulin - chicke	3.89e-05
39	128	11.7	342	2	A46743	lumican precursor - b	7.49e-05
40	128	11.7	382	2	I39068	proline- arginine-ric	7.49e-05
41	128	11.7	1115	2	S40241	G protein-coupled rec	7.49e-05
42	127	11.6	361	2	A53860	chondroadherin precur	1.04e-04
43	127	11.6	603	2	JC6128	insulin-like growth f	1.04e-04
44	124	11.4	605	2	JC5239	insulin-like growth f	2.74e-04
45	123	11.3	536	2	A34901	lysine carboxypeptida	3.77e-04

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           44.2%;  Score 483;  DB 2;  Length 1469;
  Best Local Similarity  44.2%;  Pred. No. 6.23e-67;
  Matches   69;  Conservative   40;  Mismatches  44;  Indels   3;  Gaps   3;

Db    451 KNPFICDCNLRWLAD-YLHKNPIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKL 509
          :|| ||||||:|||:  |:|| |||||||| |||:::::: :|  :|||| :|  : :
Qy      1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSESFVSM 59

Db    510 SGE-CRMDSDCPAMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFG 568
           :: | :|| ||: | | ||||||  | |: ||  ||  :|:|||: |::: :  :: :
Qy     60 YADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIH 119

Db    569 RLPHLVKLELKRNQLTGIEPNAFEGASHIQELQLGE 604
          | :|  |:|  |::| |: ::||  | ::|| | :
Qy    120 VLENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155


ESULT   2
TRY             A36665     #type complete
TLE             slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
   #authors     Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
   #journal     Genes Dev. (1990) 4:2169-2187
   #title       slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
   #cross-references MUID:91099665
   #accession   A36665
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-1480 ##label ROT
      ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
   #gene        FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
   66-91        #domain proteoglycan amino-terminal homology #label
                  PAH1\
   101-124      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
   125-148      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
   149-172      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
   173-196      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
   197-220      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
   228-272      #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
   288-313      #domain proteoglycan amino-terminal homology #label
                  PAH2\
   323-346      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
   347-370      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
   371-394      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494      #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537      #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695      #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733      #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   791-814      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
   815-838      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
   846-890      #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061    #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           44.2%;  Score 483;  DB 2;  Length 1480;
  Best Local Similarity  44.2%;  Pred. No. 6.23e-67;
  Matches   69;  Conservative   40;  Mismatches  44;  Indels   3;  Gaps   3;

Db    451 KNPFICDCNLRWLAD-YLHKNPIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKL 509
          :|| ||||||:|||:  |:|| |||||||| |||:::::: :|  :|||| :|  : :
Qy      1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSESFVSM 59

Db    510 SGE-CRMDSDCPAMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFG 568
           :: | :|| ||: | | ||||||  | |: ||  ||  :|:|||: |::: :  :: :
Qy     60 YADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIH 119

Db    569 RLPHLVKLELKRNQLTGIEPNAFEGASHIQELQLGE 604
```

```
          | :|  |:|  |::| |: ::||  | ::|| | :
Qy    120 VLENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155


RESULT    3
ENTRY           A53531     #type complete
TITLE           oncofetal trophoblast glycoprotein 5T4 precursor - human
ALTERNATE_NAMES oncofetal antigen 5T4 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change
                  10-Sep-1997
ACCESSIONS      A53531; S40087
REFERENCE       A53531
   #authors     Myers, K.A.; Rahi-Saund, V.; Davison, M.D.; Young, J.A.;
                  Cheater, A.J.; Stern, P.L.
   #journal     J. Biol. Chem. (1994) 269:9319-9324
   #title       Isolation of a cDNA encoding 5T4 oncofetal trophoblast
                  glycoprotein. An antigen associated with metastasis
                  contains leucine-rich repeats.
   #cross-references MUID:94179356
   #accession   A53531
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-420 ##label MEY
      ##cross-references EMBL:Z29083; NID:g435654; PID:g435655
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        duplication; glycoprotein; transmembrane protein
FEATURE
   1-31             #domain signal sequence #status predicted #label SIG\
   32-420           #product oncofetal trophoblast glycoprotein 5T4 #status
                      predicted #label MAT
SUMMARY         #length 420  #molecular-weight 46031  #checksum 8580

  Query Match          18.4%;  Score 201;  DB 2;  Length 420;
  Best Local Similarity  31.5%;  Pred. No. 3.72e-16;
  Matches   28;  Conservative   25;  Mismatches  32;  Indels   4;  Gaps   4;

Db     62 CPALCECSEAARTVKCVNRNLTEVPTDLPAYVRNLFLTGNQLAVLPAGAFARRPPLAELA 121
          ||: |:|  :: || | :| |  :|| :| :   |:|:|| :: :   :   :  | :|
Qy     70 CPTQCDCY-GT-TVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNS-NIHV-LENLE 125

Db    122 ALNLSGSRLDEVRAGAFEHLPSLRQLDLS 150
          |:|| :::  :   :|| |: ||:| |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    4
ENTRY           S20811     #type complete
TITLE           proteoglycan I - mouse
ALTERNATE_NAMES biglycan
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  05-Dec-1997
ACCESSIONS      S20811; A57645; I49534
REFERENCE       S20811
   #authors     Naitoh, Y.; Suzuki, S.
   #submission  submitted to the EMBL Data Library, July 1990
   #description Nucleotide sequences of cDNAs encoding mouse PGI and PGII.
   #accession   S20811
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-369 ##label NAI
      ##cross-references EMBL:X53928; NID:g53666; PID:g53667
REFERENCE       A57645
   #authors     Wegrowski, Y.; Pillarisetti, J.; Danielson, K.G.; Suzuki, S.;
                  Iozzo, R.V.
   #journal     Genomics (1995) 30:8-17
   #title       The murine biglycan: complete cDNA cloning, genomic
                  organization, promoter function, and expression.
   #cross-references MUID:96129295
   #accession   A57645
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-67,'W',69-369 ##label WEG
      ##cross-references GB:L20276; NID:g348961; PID:g348962
      ##note         authors translated the codon TGG for residue 58 as Cys
REFERENCE       I49534
   #authors     Rau, W.; Just, W.; Vetter, U.; Vogel, W.
   #journal     Mamm. Genome (1994) 5:395-396
   #title       A dinucleotide repeat in the mouse biglycan gene (EST) on the
                  X chromosome.
   #cross-references MUID:94319093
   #accession   I49534
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-67,'W',69-369 ##label RES
      ##cross-references GB:L20276; NID:g348961; PID:g348962
GENETICS
   #gene        Bgn
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        chondroitin sulfate proteoglycan; dermatan sulfate;
                  extracellular matrix; glycoprotein
FEATURE
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH
SUMMARY         #length 369  #molecular-weight 41639  #checksum 3586

  Query Match          18.0%;  Score 197;  DB 2;  Length 369;
  Best Local Similarity  34.8%;  Pred. No. 1.69e-15;
  Matches   31;  Conservative   23;  Mismatches  34;  Indels   1;  Gaps   1;

Db     60 FSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLY 118
          : ::||  | |    |:|:  ||:|:|  |:  :| | |  |:|| :  : ::  |::|
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    119 ALVLVNNKISKIHEKAFSPLRKLQKLYIS 147
          | | || |: |::|:|  | ||: | ::
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    5
ENTRY           S32793     #type complete
TITLE           biglycan precursor - rat
ALTERNATE_NAMES dermatan sulfate proteoglycan I (DS-PGI); proteoglycan I core
                  protein (PG-I)
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change
                  29-Jan-1999
ACCESSIONS      S32793
REFERENCE       S32793
   #authors     Dreher, K.L.; Asundi, V.; Matzura, D.; Cowan, K.
   #journal     Eur. J. Cell Biol. (1990) 53:296-304
```

```
   #title        Vascular smooth muscle biglycan represents a highly conserved
                   proteoglycan within the arterial wall.
   #cross-references MUID:91184222
   #accession    S32793
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-369 ##label DRE
      ##cross-references GB:U17834; NID:g600497; PID:g600498
CLASSIFICATION   #superfamily decorin; leucine-rich alpha-2-glycoprotein
                   repeat homology; proteoglycan amino-terminal homology;
                   proteoglycan carboxyl-terminal homology
KEYWORDS         chondroitin sulfate proteoglycan; dermatan sulfate;
                   extracellular matrix; glycoprotein
FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-37            #domain propeptide #status predicted #label PRO\
   38-369           #product biglycan #status predicted #label MAT\
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH\
   42,48,181,199    #binding_site dermatan sulfate (Ser) (covalent) #status
                      predicted\
   271,312          #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY          #length 369  #molecular-weight 41706  #checksum 3056

  Query Match          18.0%;  Score 197;  DB 2;  Length 369;
  Best Local Similarity  34.8%;  Pred. No. 1.69e-15;
  Matches   31;  Conservative   23;  Mismatches  34;  Indels   1;  Gaps   1;

Qy      60 FSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLY 118
           : ::||  | |    |:|:  ||:|:|  |:  :| | |  |:|| :  : ::  |::|
Db     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Qy     119 ALVLVNNKISKIHEKAFSPLRKLQKLYIS 147
           | | || |: |::|:|  | ||: | ::
Db     126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    6
ENTRY            S06280     #type complete
TITLE            decorin precursor - bovine
ALTERNATE_NAMES  dermatan sulfate proteoglycan II; proteoglycan core protein
                   II
ORGANISM         #formal_name Bos primigenius taurus #common_name cattle
DATE             31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
                   14-Nov-1997
ACCESSIONS       S06280; B31430; A26545; A20935
REFERENCE        S06280
   #authors      Day, A.A.; McQuillan, C.I.; Termine, J.D.; Young, M.R.
   #journal      Biochem. J. (1987) 248:801-805
   #title        Molecular cloning and sequence analysis of the cDNA for small
                   proteoglycan II of bovine bone.
   #cross-references MUID:88133946
   #accession    S06280
      ##molecule_type mRNA
      ##residues     1-360 ##label DAY
      ##cross-references EMBL:Y00712; NID:g618; PID:g619
      ##experimental_source bone
REFERENCE        A31430
   #authors      Choi, H.U.; Johnson, T.L.; Pal, S.; Tang, L.H.; Rosenberg,
                   L.; Neame, P.J.
   #journal      J. Biol. Chem. (1989) 264:2876-2884
   #title        Characterization of the dermatan sulfate proteoglycans,
                   DS-PGI and DS-PGII, from bovine articular cartilage and
                   skin isolated by octyl-sepharose chromatography.
   #cross-references MUID:89123388
   #accession    B31430
      ##molecule_type protein
      ##residues     31-33,'X',35-54 ##label CHO
      ##experimental_source cartilage; fetal skin
REFERENCE        A26545
   #authors      Coster, L.; Rosenberg, L.C.; van der Rest, M.; Poole, A.R.
   #journal      J. Biol. Chem. (1987) 262:3809-3812
   #title        The dermatan sulfate proteoglycans of bovine sclera and their
                   relationship to those of articular cartilage. An
                   immunological and biochemical study.
   #cross-references MUID:87137687
   #accession    A26545
      ##molecule_type protein
      ##residues     31-50 ##label COS
      ##experimental_source sclera
REFERENCE        A20935
   #authors      Pearson, C.H.; Winterbottom, N.; Fackre, D.S.; Scott, P.G.;
                   Carpenter, M.R.
   #journal      J. Biol. Chem. (1983) 258:15101-15104
   #cross-references MUID:84087911
   #accession    A20935
      ##molecule_type protein
      ##residues     31-54 ##label PEA
      ##experimental_source skin
REFERENCE        A44700
   #authors      Chopra, R.K.; Pearson, C.H.; Pringle, G.A.; Fackre, D.S.;
                   Scott, P.G.
   #journal      Biochem. J. (1985) 232:277-279
   #title        Dermatan sulphate is located on serine-4 of bovine skin
                   proteodermatan sulphate. Demonstration that most molecules
                   possess only one glycosaminoglycan chain and comparison of
                   amino acid sequences around glycosylation sites in
                   different proteoglycans.
   #contents     annotation; glycosylation
CLASSIFICATION   #superfamily decorin; leucine-rich alpha-2-glycoprotein
                   repeat homology; proteoglycan amino-terminal homology;
                   proteoglycan carboxyl-terminal homology
KEYWORDS         chondroitin sulfate proteoglycan; collagen binding; dermatan
                   sulfate; extracellular matrix; glycoprotein
FEATURE
   1-15             #domain signal sequence #status predicted #label SIG\
   16-30            #domain propeptide #status predicted #label PRO\
   31-360           #product decorin #status predicted #label MAT\
   49-73            #domain proteoglycan amino-terminal homology #label PAH\
   83-106           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   107-130          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   131-151          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   152-175          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   176-199          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   202-222          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   223-246          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   247-270          #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LRR8\
   271-293        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   294-308        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #status atypical #label LR10\
   309-360        #domain proteoglycan carboxyl-terminal homology #label
                  PCH\
   34             #binding_site dermatan sulfate (Ser) (covalent) #status
                  experimental\
   190,326        #binding_site dermatan sulfate (Ser) (covalent) #status
                  predicted\
   212,263,304    #binding_site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY         #length 360  #molecular-weight 39837  #checksum 9778

  Query Match           17.9%;  Score 195;  DB 2;  Length 360;
  Best Local Similarity  34.4%;  Pred. No. 3.59e-15;
  Matches   31;  Conservative   21;  Mismatches  35;  Indels   3;  Gaps   3;

Db     51 MGPVCPFRCQCHLRVVQCSDLGLEKVPKDLPP-DTALLDLQNNKITEIK-DGDFKNLKNL 108
          : ::|| :|:|    |:|:  ||: :|  :|   | || |  |:|: :  ::::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLL-LSGNNISTVDLNSNIHVLENL 124

Db    109 HTLILINNKISKISPGAFAPLVKLERLYLS 138
            | | || |: |:  :|  | ||  | |:
Qy    125 EXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    7
ENTRY           I47020     #type complete
TITLE           decorin - rabbit
ORGANISM        #formal_name Oryctolagus cuniculus #common_name domestic
                  rabbit
DATE            04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change
                  16-Dec-1998
ACCESSIONS      I47020
REFERENCE       I47020
   #authors     Zhan, Q.; Burrows, R.; Cintron, C.
   #journal     Invest. Ophthalmol. Vis. Sci. (1995) 36:206-215
   #title       Cloning and in situ hybridization of rabbit decorin in
                  corneal tissues.
   #cross-references MUID:95122319
   #accession   I47020
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-360 ##label ZHA
      ##cross-references GB:S76584; NID:g913374; PID:g913375
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology

FEATURE
   49-73           #domain proteoglycan amino-terminal homology #label PAH\
   83-106          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR1\
   107-130         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR2\
   131-151         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR3\
   152-175         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR4\
   176-199         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR5\
   202-222         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR6\
   223-246         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR7\
   247-270         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR8\
   271-293         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR9\
   294-308         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #status atypical #label LR10\
   309-360         #domain proteoglycan carboxyl-terminal homology #label
                     PCH
SUMMARY         #length 360  #molecular-weight 39896  #checksum 8087

  Query Match           17.9%;  Score 195;  DB 2;  Length 360;
  Best Local Similarity  33.7%;  Pred. No. 3.59e-15;
  Matches   30;  Conservative   22;  Mismatches  36;  Indels   1;  Gaps   1;

Db     51 LGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNLH 109
          : ::|| :|:|    |:|:  ||: :|  :|  :| | | |:|: :  ::::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    110 ALILVNNKISKISPGAFTPLVKLERLYLS 138
            | | || |: |:  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    8
ENTRY           NBHUC8     #type complete
TITLE           decorin precursor - human
ALTERNATE_NAMES  cartilage proteoglycan protein II; DS-PG II; PG40 core
                  protein; proteoglycan 38K core protein
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
                  31-Dec-1998
ACCESSIONS      A45016; A45015; B45015; A26476; S05640
REFERENCE       A45016
   #authors     Vetter, U.; Vogel, W.; Just, W.; Young, M.F.; Fisher, L.W.
   #journal     Genomics (1993) 15:161-168
   #title       Human decorin gene: intron-exon junctions and chromosomal
                  localization.
   #cross-references MUID:93162643
   #accession   A45016
      ##molecule_type DNA
      ##residues     1-359 ##label VET
      ##cross-references GB:L01125; GB:L01126; GB:L01127; GB:L01128;
                        GB:L01129; GB:L01130; GB:L01131
      ##note        sequence extracted from NCBI backbone (NCBIP:125061)
REFERENCE       A45015
   #authors     Danielson, K.G.; Fazzio, A.; Cohen, I.; Cannizzaro, L.A.;
                  Eichstetter, I.; Iozzo, R.V.
   #journal     Genomics (1993) 15:146-160
   #title       The human decorin gene: intron-exon organization, discovery
                  of two alternatively spliced exons in the 5' untranslated
                  region, and mapping of the gene to chromosome 12q23.
   #cross-references MUID:93162642
   #accession   A45015
      ##status       not compared with conceptual translation
      ##molecule_type DNA
      ##residues     28-70 ##label DA2
      ##cross-references GB:M98262
      ##note        sequence extracted from NCBI backbone (NCBIP:125013)
   #accession   B45015
      ##status       not compared with conceptual translation
      ##molecule_type DNA
      ##residues     296-359 ##label DAN
      ##note        sequence extracted from NCBI backbone (NCBIP:125017)
REFERENCE       A26476
   #authors     Krusius, T.; Ruoslahti, E.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7683-7687
   #title       Primary structure of an extracellular matrix proteoglycan
                  core protein deduced from cloned cDNA.
   #cross-references MUID:87017013
   #accession   A26476
      ##molecule_type mRNA
      ##residues     1-359 ##label KRU
      ##cross-references GB:M14219; NID:g181169; PID:g181170
REFERENCE       S05639
   #authors     Roughley, P.J.; White, R.J.
   #journal     Biochem. J. (1989) 262:823-827
   #title       Dermatan sulphate proteoglycans of human articular cartilage.
                  The properties of dermatan sulphate proteoglycans I and II.
   #cross-references MUID:90073579
   #accession   S05640
```

```
     ##molecule_type protein
     ##residues      31-33,'X',35-50 ##label ROU
COMMENT    This protein binds type I collagen.
GENETICS
   #gene         GDB:DCN
     ##cross-references GDB:119839; OMIM:125255
   #map_position 12q21.3-12q23
   #introns      71/1; 108/3; 180/1; 218/1; 249/2; 295/3
   #note         the first two introns occur before the initiator codon
CLASSIFICATION   #superfamily decorin; leucine-rich alpha-2-glycoprotein
                   repeat homology; proteoglycan amino-terminal homology;
                   proteoglycan carboxyl-terminal homology
KEYWORDS         chondroitin sulfate proteoglycan; collagen binding; dermatan
                   sulfate; duplication; extracellular matrix; fibroblast;
                   glycoprotein; tandem repeat
FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-30            #domain propeptide #status predicted #label PRO\
   31-359           #product decorin #status predicted #label MPT\
   48-72            #domain proteoglycan amino-terminal homology #label PAH\
   82-105           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   106-129          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   130-150          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   151-174          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   175-200          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   201-221          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   222-245          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   246-269          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   270-292          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   293-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   34               #binding_site dermatan sulfate (Ser) (covalent) #status
                      experimental\
   189,325          #binding_site dermatan sulfate (Ser) (covalent) #status
                      predicted\
   211,262,303      #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY          #length 359  #molecular-weight 39746  #checksum 835

  Query Match           17.7%;  Score 193;  DB 1;  Length 359;
  Best Local Similarity  32.6%;  Pred. No. 7.6le-15;
  Matches    29;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db     50 LGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNLH 108
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::: | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    109 ALILVNNKISKVSPGAFTPLVKLERLYLS 137
          | | || |: ::  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    9
ENTRY            BGHUN      #type complete
TITLE            biglycan precursor - human
ALTERNATE_NAMES  cartilage proteoglycan protein I; dermatan sulfate
                   proteoglycan I (DS-PGI); proteoglycan I core protein (PG-I)
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             21-Apr-1992 #sequence_revision 26-May-1995 #text_change
                   14-Nov-1997
ACCESSIONS       A40757; A32458; S05639; A28457; S14349
REFERENCE        A40757
   #authors      Fisher, L.W.; Heegaard, A.M.; Vetter, U.; Vogel, W.; Just,
                   W.; Termine, J.D.; Young, M.F.
   #journal      J. Biol. Chem. (1991) 266:14371-14377
   #title        Human biglycan gene. Putative promoter, intron-exon
                   junctions, and chromosomal localization.
   #cross-references MUID:91317791
   #accession    A40757
     ##molecule_type DNA
     ##residues      1-368 ##label FIS
     ##cross-references GB:M65151; GB:M65152; GB:M65153
REFERENCE        A32458
   #authors      Fisher, L.W.; Termine, J.D.; Young, M.F.
   #journal      J. Biol. Chem. (1989) 264:4571-4576
   #title        Deduced protein sequence of bone small proteoglycan I
                   (Biglycan) shows homology with proteoglycan II (Decorin)
                   and several nonconnective tissue proteins in a variety of
                   species.
   #cross-references MUID:89174714
   #accession    A32458
     ##molecule_type mRNA
     ##residues      1-138,'NV',141-162,'DV',165-368 ##label FI2
     ##cross-references GB:J04599
     ##note          parts of this sequence, including the amino end of the
                      mature protein, were determined by protein sequencing
REFERENCE        S05639
   #authors      Roughley, P.J.; White, R.J.
   #journal      Biochem. J. (1989) 262:823-827
   #title        Dermatan sulphate proteoglycans of human articular cartilage.
                   The properties of dermatan sulphate proteoglycans I and II.
   #cross-references MUID:90073579
   #accession    S05639
     ##molecule_type protein
     ##residues      38-41,'X',43-46,'X',48-57 ##label ROU
REFERENCE        A92656
   #authors      Fisher, L.W.; Hawkins, G.R.; Tuross, N.; Termine, J.D.
   #journal      J. Biol. Chem. (1987) 262:9702-9708
   #title        Purification and partial characterization of small
                   proteoglycans I and II, bone sialoproteins I and II, and
                   osteonectin from the mineral compartment of developing
                   human bone.
   #cross-references MUID:87250639
   #accession    A28457
     ##molecule_type protein
     ##residues      38-41,'X',43-62,'X',64-66 ##label FI3
     ##experimental_source bone
REFERENCE        S14349
   #authors      Stoecker, G.; Meyer, H.E.; Wagener, C.; Greiling, H.
   #journal      Biochem. J. (1991) 274:415-420
   #title        Purification and N-terminal amino acid sequence of a
                   chondroitin sulphate/dermatan sulphate proteoglycan
                   isolated from intima/media preparations of human aorta.
   #cross-references MUID:91174749
   #accession    S14349
     ##molecule_type protein
     ##residues      38-57 ##label STO
     ##experimental_source aorta
GENETICS
   #gene         GDB:BGN
     ##cross-references GDB:119727; OMIM:301870
   #map_position Xq28-Xq28
   #introns      80/1; 117/3; 189/1; 226/1; 257/2; 303/3
CLASSIFICATION   #superfamily decorin; leucine-rich alpha-2-glycoprotein
                   repeat homology; proteoglycan amino-terminal homology;
                   proteoglycan carboxyl-terminal homology
KEYWORDS         chondroitin sulfate proteoglycan; dermatan sulfate;
                   duplication; extracellular matrix; glycoprotein; tandem
                   repeat
FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-37            #domain propeptide #status predicted #label PRO\
   38-368           #product biglycan #status predicted #label MAT\
   57-81            #domain proteoglycan amino-terminal homology #label PAH\
   91-114           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   115-138          #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LRR2\
  139-159         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
  160-183         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
  184-207         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
  209-229         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
  230-253         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
  254-277         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  278-300         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  301-315         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #status atypical #label LR10\
  316-368         #domain proteoglycan carboxyl-terminal homology #label
                  PCH\
  42,47           #binding_site dermatan sulfate (Ser) (covalent) #status
                  experimental\
  180,198         #binding_site dermatan sulfate (Ser) (covalent) #status
                  predicted\
  270,311         #binding_site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY           #length 368  #molecular-weight 41654  #checksum 1684

  Query Match          17.7%;  Score 193;  DB 1;  Length 368;
  Best Local Similarity  34.5%;  Pred. No. 7.61e-15;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     61 AMCPFGCHCHLRVVQCSDLGLKSVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLYAL 119
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    120 VLVNNKISKIHEKAFSPLRKLQKLYIS 146
            | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  10
ENTRY             S32559     #type complete
TITLE             biglycan precursor - bovine
ALTERNATE_NAMES   dermatan sulfate proteoglycan I (DS-PGI); proteochondroitin
                    core protein; proteoglycan I core protein (PG-I)
ORGANISM          #formal_name Bos primigenius taurus #common_name cattle
DATE              03-May-1994 #sequence_revision 20-Feb-1995 #text_change
                    14-Nov-1997
ACCESSIONS        S32559; S34229; A33701; A31430; PT0078; S55673; A33137
REFERENCE         S32559
   #authors       Torok, M.A.; Evans, S.A.S.; Marcum, J.A.
   #journal       Biochim. Biophys. Acta (1993) 1173:81-84
   #title         cDNA sequence for bovine biglycan (PGI) protein core.
   #accession     S32559
      ##molecule_type mRNA
      ##residues      1-369 ##label TOR
      ##cross-references EMBL:L07953; NID:g162746
      ##experimental_source aortic smooth muscle
REFERENCE         S34229
   #authors       Marcum, J.A.; Torok, M.; Evans, S.
   #submission    submitted to the EMBL Data Library, December 1992
   #accession     S34229
      ##molecule_type mRNA
      ##residues      1-250,'V',252-369 ##label MAR
      ##cross-references EMBL:L07953
REFERENCE         A33701
   #authors       Neame, P.J.; Choi, H.U.; Rosenberg, L.C.
   #journal       J. Biol. Chem. (1989) 264:8653-8661
   #title         The primary structure of the core protein of the small,
                    leucine-rich proteoglycan (PG I) from bovine articular
                    cartilage.
   #cross-references MUID:89255324
   #accession     A33701
      ##molecule_type protein
      ##residues      38-187,'E',189-367,'Y' ##label NEA
      ##experimental_source cartilage
REFERENCE         A31430
   #authors       Choi, H.U.; Johnson, T.L.; Pal, S.; Tang, L.H.; Rosenberg,
                    L.; Neame, P.J.
   #journal       J. Biol. Chem. (1989) 264:2876-2884
   #title         Characterization of the dermatan sulfate proteoglycans,
                    DS-PGI and DS-PGII, from bovine articular cartilage and
                    skin isolated by octyl-sepharose chromatography.
   #cross-references MUID:89123388
   #accession     A31430
      ##molecule_type protein
      ##residues      38-41,'X',43-47,'X',49-63 ##label CHO
      ##note          sequences from skin and cartilage were identical
REFERENCE         PT0077
   #authors       Marcum, J.A.; Thompson, M.A.
   #journal       Biochem. Biophys. Res. Commun. (1991) 175:706-712
   #title         The amino-terminal region of a proteochondroitin core
                    protein, secreted by aortic smooth muscle cells, shares
                    sequence homology with the pre-propeptide region of the
                    biglycan core protein from human bone.
   #cross-references MUID:91207372
   #accession     PT0078
      ##molecule_type protein
      ##residues      17-24,'F',26-30 ##label MA2
      ##experimental_source aortic smooth muscle
REFERENCE         S55673
   #authors       Scott, P.G.; Nakano, T.; Dodd, C.M.
   #journal       Biochim. Biophys. Acta (1995) 1244:121-128
   #title         Small proteoglycans from different regions of the
                    fibrocartilaginous temporomandibular joint disc.
   #cross-references MUID:95284073
   #accession     S55673
      ##molecule_type protein
      ##residues      38-41,'X',43-47,'X',49-53 ##label SCO
CLASSIFICATION    #superfamily decorin; leucine-rich alpha-2-glycoprotein
                    repeat homology; proteoglycan amino-terminal homology;
                    proteoglycan carboxyl-terminal homology
KEYWORDS          cartilage; chondroitin sulfate proteoglycan; dermatan
                    sulfate; extracellular matrix; glycoprotein

FEATURE
   1-16             #domain signal sequence #status predicted #label SIG\
   17-37            #domain amino-terminal propeptide #status predicted
                      #label PRO\
   38-369           #product biglycan #status predicted #label MAT\
   58-82            #domain proteoglycan amino-terminal homology #label PAH\
   92-115           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   116-139          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   140-160          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   161-184          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   185-208          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   210-230          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   231-254          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   255-278          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   279-301          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   302-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
   317-369          #domain proteoglycan carboxyl-terminal homology #label
                      PCH\
   42,48            #binding_site dermatan sulfate (Ser) (covalent) #status
                      experimental\
   181,199          #binding_site dermatan sulfate (Ser) (covalent) #status
                      predicted\
```

```
  271,312          #binding_site carbohydrate (Asn) (covalent) #status
                     predicted
SUMMARY          #length 369  #molecular-weight 41590  #checksum 1525

  Query Match           17.7%;  Score 193;  DB 2;  Length 369;
  Best Local Similarity  34.5%;  Pred. No. 7.61e-15;
  Matches    30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     62 AMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLYAL 120
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    121 VLVNNKISKIHEKAFSPLRKLQKLYIS 147
          | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   11
ENTRY            S24317     #type complete
TITLE            decorin precursor - chicken
ALTERNATE_NAMES  corneal chondroitin/dermatan sulfate proteoglycan
ORGANISM         #formal_name Gallus gallus #common_name chicken
DATE             13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                   14-Nov-1997
ACCESSIONS       S24317; S58474; S22197
REFERENCE        S24317
   #authors      Li, W.; Vergnes, J.P.; Cornuet, P.K.; Hassell, J.R.
   #journal      Arch. Biochem. Biophys. (1992) 296:190-197
   #title        cDNA clone to chick corneal chondroitin/dermatan sulfate
                   proteoglycan reveals identity to decorin.
   #cross-references MUID:92296755
   #accession    S24317
      ##molecule_type mRNA
      ##residues      1-357 ##label LIW
      ##cross-references EMBL:X63797; NID:g62887; PID:g62888
   #accession    S58474
      ##molecule_type protein
      ##residues      31-33,'X',35-39,'X',41-48,'X',50-51 ##label LIA
CLASSIFICATION   #superfamily decorin; leucine-rich alpha-2-glycoprotein
                   repeat homology; proteoglycan amino-terminal homology;
                   proteoglycan carboxyl-terminal homology
KEYWORDS         collagen binding; extracellular matrix; glycoprotein
FEATURE
   1-16              #domain signal sequence #status predicted #label SIG\
   17-30             #domain propeptide #status predicted #label PRO\
   31-357            #product decorin #status experimental #label MAT\
   46-70             #domain proteoglycan amino-terminal homology #label PAH\
   80-103            #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   104-127           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   128-148           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   199-219           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6\
   220-243           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR7\
   244-267           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR8\
   268-290           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR9\
   291-305           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #status atypical #label LR10\
   306-357           #domain proteoglycan carboxyl-terminal homology #label
                       PCH
SUMMARY          #length 357  #molecular-weight 39687  #checksum 4375

  Query Match           17.6%;  Score 192;  DB 2;  Length 357;
  Best Local Similarity  33.7%;  Pred. No. 1.11e-14;
  Matches    30;  Conservative   22;  Mismatches  36;  Indels   1;  Gaps   1;

Db     48 FGPVCPFRCQCHLRVVQCSDLGLERVPKDLPPDTTLLDLQNNKITEIK-EGDFKNLKNLH 106
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    107 ALILVNNKISKISPAAFAPLKKLERLYLS 135
          | | || |: |:  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   12
ENTRY            S72271     #type complete
TITLE            proteoglycan Lb precursor - mouse
ALTERNATE_NAMES  chondroitin/dermatan sulphate proteoglycan
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change
                   13-Sep-1998
ACCESSIONS       S72271
REFERENCE        S72271
   #authors      Kurita, K.; Shinomura, T.; Ujita, M.; Zako, M.; Kida, D.;
                   Iwata, H.; Kimata, K.
   #journal      Biochem. J. (1996) 318:909-914
   #title        Occurrence of PG-Lb, a leucine-rich small
                   chondroitin/dermatan sulphate proteoglycan in mammalian
                   epiphyseal cartilage: molecular cloning and sequence
                   analysis of the mouse cDNA.
   #accession    S72271
      ##molecule_type mRNA
      ##residues      1-322 ##label KUR
      ##cross-references EMBL:D78274; NID:g1620004; PID:d1011999; PID:g1620005
      ##experimental_source strain BALB/c; newborne; epiphyseal cartilage
FUNCTION
   #description  probably participates in ossification process
CLASSIFICATION   #superfamily osteoinductive factor precursor; leucine-rich
                   alpha-2-glycoprotein repeat homology
KEYWORDS         glycoprotein
FEATURE
   1-25              #domain signal sequence #status predicted #label SIG\
   26-322            #product proteoglycan Lb #status predicted #label MAT\
   145-167           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   168-191           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   192-211           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   212-237           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   238-258           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   259-282           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6\
   283,302           #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 322  #molecular-weight 36762  #checksum 9522

  Query Match           16.8%;  Score 184;  DB 2;  Length 322;
  Best Local Similarity  34.1%;  Pred. No. 2.18e-13;
  Matches    28;  Conservative   27;  Mismatches  25;  Indels   2;  Gaps   2;

Db    121 CTCISTTVYCDDHELDAIP-PLPKKTTYFYSRFNRIKKIN-KNDFASLNDLKRIDLTSNL 178
          | | :||| |: : |::|| ::|: :| :    | |  :: ::::  |::|  :||::|
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXLDLSNNH 133

Db    179 ISEIDEDAFRKLPHLQELVLRD 200
          |: |:: :| ||: |:|| | |
Qy    134 ITFINDKSFEKLSKLRELXLND 155


RESULT   13
ENTRY            A55454     #type complete
TITLE            decorin precursor - mouse
ALTERNATE_NAMES  proteoglycan II
```

```
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change
                  14-Nov-1997
ACCESSIONS      A55454; S20812
REFERENCE       A55454
  #authors      Scholzen, T.; Solursh, M.; Suzuki, S.; Reiter, R.; Morgan,
                  J.L.; Buchberg, A.M.; Siracusa, L.D.; Iozzo, R.V.
  #journal      J. Biol. Chem. (1994) 269:28270-28281
  #title        The murine decorin. Complete cDNA cloning, genomic
                  organization, chromosomal assignment, and expression during
                  organogenesis and tissue differentiation.
  #cross-references MUID:95050610
  #accession    A55454
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-354 ##label SCH
    ##cross-references GB:X53929; NID:g53668; PID:g53669
REFERENCE       S20811
  #authors      Naitoh, Y.; Suzuki, S.
  #submission   submitted to the EMBL Data Library, July 1990
  #description  Nucleotide sequences of cDNAs encoding mouse PGI and PGII.
  #accession    S20812
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-354 ##label NAI
    ##cross-references EMBL:X53929; NID:g53668; PID:g53669
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        collagen binding; extracellular matrix; glycoprotein
FEATURE
  1-16              #domain signal sequence #status predicted #label SIG\
  17-30             #domain propeptide #status predicted #label PRO\
  43-67             #domain proteoglycan amino-terminal homology #label PAH\
  77-100            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  125-145           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  146-169           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  170-193           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  196-216           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  217-240           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  241-264           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
  265-287           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
  288-302           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
  303-354           #domain proteoglycan carboxyl-terminal homology #label
                      PCH
SUMMARY         #length 354  #molecular-weight 39809  #checksum 3271

  Query Match          16.8%;  Score 184;  DB 2;  Length 354;
  Best Local Similarity  34.4%;  Pred. No. 2.18e-13;
  Matches   33;  Conservative   23;  Mismatches  37;  Indels   3;  Gaps   3;

Db     39 YDPDNPLI-SMCPYRCQCHLRVVQCSDLGLDKVPWDFPPDTTLLDLQNNKITEIK-EGAF 96
          | :|: :| |:|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :: :
Qy     60 Y-ADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNI 118

Db     97 KNLKDLHTLILVNNKISKISPEAFKPLVKLERLYLS 132
             | :|  | | || |: |:  :|  | ||  | |:
Qy    119 HVLENLEXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   14
ENTRY           S29145     #type complete
TITLE           decorin precursor - rat
ALTERNATE_NAMES dermatan sulfate proteoglycan-II
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                  17-Mar-1999
ACCESSIONS      S29145; I60238; S28517
REFERENCE       S29145
  #authors      Abramson, S.R.; Woessner Jr., J.F.
  #journal      Biochim. Biophys. Acta (1992) 1132:225-227
  #title        cDNA sequence for rat dermatan sulfate proteoglycan-II
                  (decorin).
  #cross-references MUID:93003331
  #accession    S29145
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-354 ##label ABR
    ##cross-references EMBL:Z12298; NID:g57549; PID:g57550
REFERENCE       I60238
  #authors      Asundi, V.K.; Dreher, K.L.
  #journal      Eur. J. Cell Biol. (1992) 59:314-321
  #title        Molecular characterization of vascular smooth muscle decorin:
                  deduced core protein structure and regulation of gene
                  expression.
  #cross-references MUID:93154359
  #accession    I60238
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues     11-354 ##label RES
    ##cross-references EMBL:X59859; NID:g56056; PID:g56057
GENETICS
  #gene         DCN
CLASSIFICATION  #superfamily decorin; leucine-rich alpha-2-glycoprotein
                  repeat homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
KEYWORDS        collagen binding; extracellular matrix; glycoprotein
FEATURE
  1-30              #domain signal sequence (fragment) #status predicted
                      #label SIG\
  31-354            #product decorin #status predicted #label MAT\
  43-67             #domain proteoglycan amino-terminal homology #label PAH\
  77-100            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  125-145           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  146-169           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  170-193           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  196-216           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  217-240           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  241-264           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
  265-287           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
  288-302           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #status atypical #label LR10\
  303-354           #domain proteoglycan carboxyl-terminal homology #label
                      PCH
SUMMARY         #length 354  #molecular-weight 39805  #checksum 4526

  Query Match          16.8%;  Score 183;  DB 2;  Length 354;
  Best Local Similarity  34.4%;  Pred. No. 3.16e-13;
  Matches   33;  Conservative   23;  Mismatches  37;  Indels   3;  Gaps   3;

Db     39 YDPDNPLI-SMCPYRCQCHLRVVQCSDLGLDKVPWEFPPDTTLLDLQNNKITEIK-EGAF 96
          | :|: :| |:|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :: :
Qy     60 Y-ADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNI 118

Db     97 KNLKDLHTLILVNNKISKISPEAFKPLVKLERLYLS 132
```

```
          | :|  | | || |: |:  :|  | ||  | |:
Qy    119 HVLENLEXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  15
ENTRY           A41781     #type complete
TITLE           proteoglycan-Lb - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                  04-Sep-1998
ACCESSIONS      A41781
REFERENCE       A41781
  #authors      Shinomura, T.; Kimata, K.
  #journal      J. Biol. Chem. (1992) 267:1265-1270
  #title        Proteoglycan-Lb, a small dermatan sulfate proteoglycan
                  expressed in embryonic chick epiphyseal cartilage, is
                  structurally related to osteoinductive factor.
  #cross-references MUID:92112771
  #accession    A41781
    ##status        preliminary
    ##molecule_type mRNA
    ##residues      1-316 ##label SHI
    ##cross-references GB:D10485; GB:D90461; NID:g222846; PID:d1001838;
                        PID:g222847
    ##experimental_source embryo
    ##note          sequence extracted from NCBI backbone (NCBIN:76536,
                    NCBIP:76537)
CLASSIFICATION  #superfamily osteoinductive factor precursor; leucine-rich
                  alpha-2-glycoprotein repeat homology
FEATURE
  139-161           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  162-185           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  186-205           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  206-231           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  232-252           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  253-276           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6
SUMMARY         #length 316  #molecular-weight 35856  #checksum 111

 Query Match           16.7%;  Score 182;  DB 2;  Length 316;
 Best Local Similarity  36.1%;  Pred. No. 4.57e-13;
 Matches   30;  Conservative   23;  Mismatches  26;  Indels   4;  Gaps   4;

Db    115 CTCLGTTVYCDDRELDAVP-PLPK-NTMYFYSRYNRIRKIN-KNDFANLNNLKRIDLTAN 171
          | | |||| |: | |:::| ::|:  |  : |  | |  :: ::::  |:||  :||: |
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSG-NNISTVDLNSNIHVLENLEXLDLSNN 132

Db    172 LISEIHEDAFRRLPQLLELVLRD 194
          |: |:: :| :|: | || | |
Qy    133 HITFINDKSFEKLSKLRELXLND 155


Search completed: Fri May 28 09:04:25 1999
Job time : 28 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:04:42 1999; MasPar time 7.04 Seconds
622.440 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-8
Description: (1-155) from US09191647.pep
Perfect Score: 1092
Sequence: 1 RNPXICDCNLQWLAQINLQK.....FINDKSFEKLSKLRELXLND 155

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 44.155; Variance 77.249; scale 0.572

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	483	44.2	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	1.19e-80
2	197	18.0	369	1	PGS1_RAT	BONE/CARTILAGE PROTEOG	4.95e-19
3	197	18.0	369	1	PGS1_MOUSE	BONE/CARTILAGE PROTEOG	4.95e-19
4	195	17.9	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	1.22e-18
5	193	17.7	359	1	PGS2_HUMAN	BONE PROTEOGLYCAN II P	2.98e-18
6	193	17.7	360	1	PGS2_CANFA	BONE PROTEOGLYCAN II P	2.98e-18
7	193	17.7	368	1	PGS1_HUMAN	BONE/CARTILAGE PROTEOG	2.98e-18
8	193	17.7	369	1	PGS1_BOVIN	BONE/CARTILAGE PROTEOG	2.98e-18
9	192	17.6	357	1	PGS2_CHICK	BONE PROTEOGLYCAN II P	4.66e-18
10	189	17.3	369	1	PGS1_CANFA	BONE/CARTILAGE PROTEOG	1.78e-17
11	184	16.8	354	1	PGS2_MOUSE	BONE PROTEOGLYCAN II P	1.63e-16
12	183	16.8	354	1	PGS2_RAT	BONE PROTEOGLYCAN II P	2.53e-16
13	162	14.8	208	1	GPBB_PAPCY	PLATELET GLYCOPROTEIN	2.24e-12
14	159	14.6	206	1	GPBB_HUMAN	PLATELET GLYCOPROTEIN	7.95e-12
15	159	14.6	206	1	GPBB_MOUSE	PLATELET GLYCOPROTEIN	7.95e-12
16	159	14.6	343	1	LUM_CHICK	LUMICAN PRECURSOR (LUM	7.95e-12
17	147	13.5	662	1	GARP_HUMAN	GARP PROTEIN PRECURSOR	1.16e-09
18	142	13.0	440	1	OMGP_MOUSE	OLIGODENDROCYTE-MYELIN	8.89e-09
19	140	12.8	338	1	LUM_MOUSE	LUMICAN PRECURSOR (LUM	1.99e-08
20	140	12.8	440	1	OMGP_HUMAN	OLIGODENDROCYTE-MYELIN	1.99e-08
21	139	12.7	560	1	GPV_HUMAN	PLATELET GLYCOPROTEIN	2.97e-08
22	137	12.5	567	1	GPV_RAT	PLATELET GLYCOPROTEIN	6.59e-08
23	136	12.5	682	1	CONN_DROME	CONNECTIN PRECURSOR.	9.81e-08
24	136	12.5	4303	1	PKD1_HUMAN	POLYCYSTIN PRECURSOR (	9.81e-08
25	134	12.3	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	2.16e-07
26	131	12.0	1134	1	CHAO_DROME	CHAOPTIN PRECURSOR (PH	7.01e-07
27	130	11.9	298	1	OIF_HUMAN	OSTEOINDUCTIVE FACTOR	1.03e-06
28	130	11.9	299	1	OIF_BOVIN	OSTEOINDUCTIVE FACTOR	1.03e-06
29	130	11.9	338	1	LUM_RAT	LUMICAN PRECURSOR (LUM	1.03e-06
30	130	11.9	380	1	FMOD_CHICK	FIBROMODULIN PRECURSOR	1.03e-06
31	128	11.7	342	1	LUM_BOVIN	LUMICAN PRECURSOR (LUM	2.24e-06
32	128	11.7	382	1	PARG_HUMAN	PROLARGIN PRECURSOR (P	2.24e-06
33	128	11.7	1115	1	GPCR_LYMST	G-PROTEIN COUPLED RECE	2.24e-06
34	127	11.6	361	1	CHAD_BOVIN	CHONDROADHERIN PRECURS	3.30e-06
35	127	11.6	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	3.30e-06
36	124	11.4	338	1	LUM_HUMAN	LUMICAN PRECURSOR (LUM	1.04e-05
37	124	11.4	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	1.04e-05
38	123	11.3	536	1	CBP8_HUMAN	CARBOXYPEPTIDASE N 83	1.52e-05
39	123	11.3	695	1	FSHR_MACFA	FOLLICLE STIMULATING H	1.52e-05
40	122	11.2	692	1	FSHR_RAT	FOLLICLE STIMULATING H	2.21e-05
41	122	11.2	695	1	FSHR_HUMAN	FOLLICLE STIMULATING H	2.21e-05
42	120	11.0	695	1	FSHR_PIG	FOLLICLE STIMULATING H	4.69e-05
43	119	10.9	695	1	FSHR_BOVIN	FOLLICLE STIMULATING H	6.82e-05
44	118	10.8	694	1	FSHR_HORSE	FOLLICLE STIMULATING H	9.88e-05
45	116	10.6	567	1	GPV_MOUSE	PLATELET GLYCOPROTEIN	2.07e-04

ALIGNMENTS

```
RESULT   1
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match           44.2%;  Score 483;  DB 1;  Length 1480;
  Best Local Similarity  44.2%;  Pred. No. 1.19e-80;
  Matches    69;  Conservative   40;  Mismatches  44;  Indels   3;  Gaps   3;

Db    451 KNPFICDCNLRWLAD-YLHKNPIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKL 509
          :|| ||||||:|||:  |:|| ||||||||| |||:::::: :|  :|||| :|  : :
Qy      1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSESFVSM 59

Db    510 SGE-CRMDSDCPAMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFG 568
           :: | :|| ||: | | |||||  | |: ||  ||  :|:|||: |::: :  :: :
Qy     60 YADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIH 119

Db    569 RLPHLVKLELKRNQLTGIEPNAFEGASHIQELQLGE 604
           | :|  |:|  |::| |: ::||  | ::|| | :
Qy    120 VLENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155


RESULT    2
ID   PGS1_RAT       STANDARD;      PRT;   369 AA.
AC   P47853;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=VASCULAR SMOOTH MUSCLE;
RX   MEDLINE; 91184222.
RA   DREHER K.L., ASUNDI V., MATZURA D., COWAN K.;
RT   "Vascular smooth muscle biglycan represents a highly conserved
RT   proteoglycan within the arterial wall.";
RL   EUR. J. CELL BIOL. 53:296-304(1990).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17834; G600498; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    19       POTENTIAL.
FT   PROPEP      20    37
FT   CHAIN       38   369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72   343       LEUCINE-RICH REPEATS.
FT   REPEAT      72    95       LRR 1.
FT   REPEAT      96   116       LRR 2.
FT   REPEAT     117   140       LRR 3.
FT   REPEAT     141   164       LRR 4.
FT   REPEAT     165   185       LRR 5.
FT   REPEAT     186   211       LRR 6.
FT   REPEAT     212   231       LRR 7.
FT   REPEAT     232   255       LRR 8.
FT   REPEAT     256   276       LRR 9.
FT   REPEAT     277   302       LRR 10.
FT   REPEAT     303   321       LRR 11.
FT   REPEAT     336   343       LRR 12.
FT   CARBOHYD    42    42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48    48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271   271       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   DISULFID    64    77       BY SIMILARITY.
FT   DISULFID   322   355       BY SIMILARITY.
SQ   SEQUENCE   369 AA;  41706 MW;  6555ECED CRC32;

  Query Match          18.0%;  Score 197;  DB 1;  Length 369;
  Best Local Similarity  34.8%;  Pred. No. 4.95e-19;
  Matches   31;  Conservative   23;  Mismatches  34;  Indels   1;  Gaps   1;

Db     60 FSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLY 118
          : ::||  | |    |:|:  ||:|:|  |:  :| | |  |:|| :  : ::  |::|
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    119 ALVLVNNKISKIHEKAFSPLRKLQKLYIS 147
           | | || |: |::|:|  | ||: | ::
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   3
     PGS1_MOUSE     STANDARD;      PRT;   369 AA.
     P28653; Q61355;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS; TISSUE=FIBROBLAST;
RA   NAITOH Y., SUZUKI S.;
RL   SUBMITTED (JUL-1990) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS; TISSUE=EMBRYO;
RX   MEDLINE; 94319093.
RA   RAU W., JUST W., VETTER U., VOGEL W.;
RT   "A dinucleotide repeat in the mouse biglycan gene (EST) on the X
RT   chromosome.";
RL   MAMM. GENOME 5:395-396(1994).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53928; G53667; -.
DR   EMBL; L20276; G348962; -.
DR   PIR; S20811; S20811.
DR   MGD; MGI:88158; BGN.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    19       POTENTIAL.
FT   PROPEP      20    37
FT   CHAIN       38   369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72   343       LEUCINE-RICH REPEATS.
FT   REPEAT      72    95       LRR 1.
FT   REPEAT      96   116       LRR 2.
FT   REPEAT     117   140       LRR 3.
FT   REPEAT     141   164       LRR 4.
FT   REPEAT     165   185       LRR 5.
FT   REPEAT     186   211       LRR 6.
FT   REPEAT     212   231       LRR 7.
FT   REPEAT     232   255       LRR 8.
FT   REPEAT     256   276       LRR 9.
FT   REPEAT     277   302       LRR 10.
FT   REPEAT     303   321       LRR 11.
FT   REPEAT     336   343       LRR 12.
FT   CARBOHYD    42    42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48    48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271   271       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   DISULFID    64    77       BY SIMILARITY.
FT   DISULFID   322   355       BY SIMILARITY.
FT   CONFLICT    68    68       C -> W (IN REF. 2).
SQ   SEQUENCE   369 AA;  41639 MW;  ED21DD6B CRC32;

  Query Match          18.0%;  Score 197;  DB 1;  Length 369;
  Best Local Similarity  34.8%;  Pred. No. 4.95e-19;
  Matches   31;  Conservative   23;  Mismatches  34;  Indels   1;  Gaps   1;

Db     60 FSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLY 118
          : ::||  | |    |:|:  ||:|:|  |:  :| | |  |:|| :  : ::  |::|
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    119 ALVLVNNKISKIHEKAFSPLRKLQKLYIS 147
           | | || |: |::|:|  | ||: | ::
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   4
ID   PGS2_BOVIN     STANDARD;      PRT;   360 AA.
AC   P21793;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN).
GN   DCN.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88133946.
RA   DAY A.A., MCQUILLAN C.I., TERMINE J.D., YOUNG M.R.;
RT   "Molecular cloning and sequence analysis of the cDNA for small
```

```
RT   proteoglycan II of bovine bone.";
RL   BIOCHEM. J. 248:801-805(1987).
RN   [2]
RP   SEQUENCE OF 31-54.
RX   MEDLINE; 89123388.
RA   CHOI H.U., JOHNSON T.L., PAL S., TANG L.H., ROSENBERG L., NEAME P.J.;
RT   "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT   DS-PGII, from bovine articular cartilage and skin isolated by octyl-
RT   sepharose chromatography.";
RL   J. BIOL. CHEM. 264:2876-2884(1989).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00712; G619; -.
DR   PIR; S06280; S06280.
DR   PIR; B31430; B31430.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30
FT   CHAIN       31   360       BONE PROTEOGLYCAN II.
FT   DOMAIN      78   309       LEUCINE-RICH REPEATS.
FT   REPEAT      78    99       LRR 1.
FT   REPEAT     100   123       LRR 2.
FT   REPEAT     124   146       LRR 3.
FT   REPEAT     147   168       LRR 4.
FT   REPEAT     169   194       LRR 5.
FT   REPEAT     195   218       LRR 6.
FT   REPEAT     219   239       LRR 7.
FT   REPEAT     240   263       LRR 8.
FT   REPEAT     264   286       LRR 9.
FT   REPEAT     287   309       LRR 10.
FT   CARBOHYD    34    34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   212   212       POTENTIAL.
FT   CARBOHYD   263   263       POTENTIAL.
FT   CARBOHYD   304   304       POTENTIAL.
FT   DISULFID    55    68       BY SIMILARITY.
FT   DISULFID   314   347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39837 MW;  92C2A1A5 CRC32;

  Query Match          17.9%;  Score 195;  DB 1;  Length 360;
  Best Local Similarity  34.4%;  Pred. No. 1.22e-18;
  Matches   31;  Conservative   21;  Mismatches  35;  Indels   3;  Gaps   3;

Db     51 MGPVCPFRCQCHLRVVQCSDLGLEKVPKDLPP-DTALLDLQNNKITEIK-DGDFKNLKNL 108
          : ::|| :|:|    |:|:  ||: :|  :|   | || |  |:|: :  :::: | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLL-LSGNNISTVDLNSNIHVLENL 124

Db    109 HTLILINNKISKISPGAFAPLVKLERLYLS 138
            | | || |: |:  :|  | ||  | |:
Qy    125 EXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    5
ID   PGS2_HUMAN     STANDARD;      PRT;   359 AA.
AC   P07585;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN) (PG40).
GN   DCN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87017013.
RA   KRUSIUS T., RUOSLAHTI E.;
RT   "Primary structure of an extracellular matrix proteoglycan core
RT   protein deduced from cloned cDNA.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 83:7683-7687(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 93162643.
RA   VETTER U., VOGEL W., JUST W., YOUNG M.F., FISHER L.W.;
RT   "Human decorin gene: intron-exon junctions and chromosomal
RT   localization.";
RL   GENOMICS 15:161-168(1993).
RN   [3]
RP   SEQUENCE OF 1-70 FROM N.A.
RX   MEDLINE; 93162642.
RA   DANIELSON K.G., FAZZIO A., COHEN I.R., CANNIZZARO L., IOZZO R.V.;
RT   "The human decorin gene: intron-exon organization, discovery of two
RT   alternatively spliced exons in the 5' untranslated region, and
RT   mapping of the gene to chromosome 12q23.";
RL   GENOMICS 15:146-160(1993).
RN   [4]
RP   SEQUENCE OF 31-50.
RX   MEDLINE; 90073579.
RA   ROUGHLEY P.J., WHITE R.J.;
RT   "Dermatan sulphate proteoglycans of human articular cartilage. The
RT   properties of dermatan sulphate proteoglycans I and II.";
RL   BIOCHEM. J. 262:823-827(1989).
RN   [5]
RP   SEQUENCE OF 31-49.
RX   MEDLINE; 87250639.
RA   FISHER L.W., HAWKINS G.R., TUROSS N., TERMINE J.D.;
RT   "Purification and partial characterization of small proteoglycans I
RT   and II, bone sialoproteins I and II, and osteonectin from the mineral
RT   compartment of developing human bone.";
RL   J. BIOL. CHEM. 262:9702-9708(1987).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14219; G181170; -.
DR   EMBL; L01131; G181519; ALT_SEQ.
DR   EMBL; L01125; G181519; JOINED.
DR   EMBL; L01126; G181519; JOINED.
DR   EMBL; L01127; G181519; JOINED.
DR   EMBL; L01129; G181519; JOINED.
DR   EMBL; L01130; G181519; JOINED.
DR   EMBL; M98262; G609452; -.
```

```
DR   PIR; A26476; NBHUC8.
DR   PIR; S05640; S05640.
DR   PIR; B28457; B28457.
DR   PIR; A45016; A45016.
DR   MIM; 125255; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30
FT   CHAIN       31   359       BONE PROTEOGLYCAN II.
FT   DOMAIN      77   308       LEUCINE-RICH REPEATS.
FT   REPEAT      77    98       LRR 1.
FT   REPEAT      99   122       LRR 2.
FT   REPEAT     123   145       LRR 3.
FT   REPEAT     146   167       LRR 4.
FT   REPEAT     168   193       LRR 5.
FT   REPEAT     194   217       LRR 6.
FT   REPEAT     218   238       LRR 7.
FT   REPEAT     239   262       LRR 8.
FT   REPEAT     263   285       LRR 9.
FT   REPEAT     286   308       LRR 10.
FT   CARBOHYD    34    34       GLYCOSAMINOGLYCAN.
FT   CARBOHYD   211   211       POTENTIAL.
FT   CARBOHYD   262   262       POTENTIAL.
FT   CARBOHYD   303   303       POTENTIAL.
FT   DISULFID    54    67       BY SIMILARITY.
FT   DISULFID   313   346       BY SIMILARITY.
FT   CONFLICT    37    37       G -> A (IN REF. 5).
FT   CONFLICT    45    45       D -> P (IN REF. 5).
SQ   SEQUENCE   359 AA;  39746 MW;  F2051F3A CRC32;

  Query Match          17.7%;  Score 193;  DB 1;  Length 359;
  Best Local Similarity  32.6%;  Pred. No. 2.98e-18;
  Matches   29;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db     50 LGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNLH 108
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  ::::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    109 ALILVNNKISKVSPGAFTPLVKLERLYLS 137
           | | || |: ::  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   6
   PGS2_CANFA     STANDARD;      PRT;   360 AA.
AC   Q29393;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN).
GN   DCN OR DCN1C.
OS   CANIS FAMILIARIS (DOG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GLANT T.T.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 244-259 FROM N.A.
RA   VENTA P.J., BROUILLETTE J.A., YUZBASIYAN-GURKAN V., BREWER G.J.;
RL   SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U83141; G1916848; -.
DR   EMBL; L77684; G1280195; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30       BY SIMILARITY.
FT   CHAIN       31   360       BONE PROTEOGLYCAN II.
FT   DOMAIN      78   309       LEUCINE-RICH REPEATS.
FT   REPEAT      78    99       LRR 1.
FT   REPEAT     100   123       LRR 2.
FT   REPEAT     124   146       LRR 3.
FT   REPEAT     147   168       LRR 4.
FT   REPEAT     169   194       LRR 5.
FT   REPEAT     195   218       LRR 6.
FT   REPEAT     219   239       LRR 7.
FT   REPEAT     240   263       LRR 8.
FT   REPEAT     264   286       LRR 9.
FT   REPEAT     287   309       LRR 10.
FT   CARBOHYD    34    34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   212   212       POTENTIAL.
FT   CARBOHYD   263   263       POTENTIAL.
FT   CARBOHYD   304   304       POTENTIAL.
FT   DISULFID    55    68       BY SIMILARITY.
FT   DISULFID   314   347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39980 MW;  601D533F CRC32;

  Query Match          17.7%;  Score 193;  DB 1;  Length 360;
  Best Local Similarity  33.3%;  Pred. No. 2.98e-18;
  Matches   30;  Conservative   23;  Mismatches  36;  Indels   1;  Gaps   1;

Db     50 LLGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNL 108
          :: ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  ::::  | ||
Qy     65 FIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENL 124

Db    109 HTLILVNNKISKISPGAFTPLLKLERLYLS 138
           | | || |: |:  :|  | ||  | |:
Qy    125 EXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   7
ID   PGS1_HUMAN     STANDARD;      PRT;   368 AA.
AC   P21810; P13247;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE;
RX   MEDLINE; 89174714.
RA   FISHER L.W., TERMINE J.D., YOUNG M.F.;
RT   "Deduced protein sequence of bone small proteoglycan I (biglycan)
RT   shows homology with proteoglycan II (decorin) and several
RT   nonconnective tissue proteins in a variety of species.";
RL   J. BIOL. CHEM. 264:4571-4576(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91317791.
```

```
RA   FISHER L.W., HEEGAARD A.M., VETTER U., VOGEL W., JUST W.,
RA   TERMINE J.D., YOUNG M.F.;
RT   "Human biglycan gene. Putative promoter, intron-exon junctions, and
RT   chromosomal localization.";
RL   J. BIOL. CHEM. 266:14371-14377(1991).
RN   [3]
RP   SEQUENCE OF 38-57.
RX   MEDLINE; 90073579.
RA   ROUGHLEY P.J., WHITE R.J.;
RT   "Dermatan sulphate proteoglycans of human articular cartilage. The
RT   properties of dermatan sulphate proteoglycans I and II.";
RL   BIOCHEM. J. 262:823-827(1989).
RN   [4]
RP   SEQUENCE OF 38-66.
RX   MEDLINE; 87250639.
RA   FISHER L.W., HAWKINS G.R., TUROSS N., TERMINE J.D.;
RT   "Purification and partial characterization of small proteoglycans I
RT   and II, bone sialoproteins I and II, and osteonectin from the mineral
RT   compartment of developing human bone.";
RL   J. BIOL. CHEM. 262:9702-9708(1987).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J04599; G306884; -.
DR   EMBL; M65153; G179433; ALT_SEQ.
DR   EMBL; M65152; G179433; JOINED.
DR   PIR; A28457; A28457.
DR   PIR; A32458; A32458.
DR   PIR; A40757; A40757.
DR   PIR; S05639; S05639.
DR   MIM; 301870; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     19       POTENTIAL.
FT   PROPEP      20     37
FT   CHAIN       38    368       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      71    342       LEUCINE-RICH REPEATS.
FT   REPEAT      71     94       LRR 1.
FT   REPEAT      95    115       LRR 2.
FT   REPEAT     116    139       LRR 3.
FT   REPEAT     140    163       LRR 4.
FT   REPEAT     164    184       LRR 5.
FT   REPEAT     185    210       LRR 6.
FT   REPEAT     211    230       LRR 7.
FT   REPEAT     231    254       LRR 8.
FT   REPEAT     255    275       LRR 9.
FT   REPEAT     276    301       LRR 10.
FT   REPEAT     302    320       LRR 11.
FT   REPEAT     335    342       LRR 12.
FT   CARBOHYD    42     42       GLYCOSAMINOGLYCAN.
FT   CARBOHYD    47     47       GLYCOSAMINOGLYCAN.
FT   CARBOHYD   270    270       POTENTIAL.
FT   CARBOHYD   311    311       POTENTIAL.
FT   DISULFID    63     76       BY SIMILARITY.
FT   DISULFID   321    354       BY SIMILARITY.
FT   CONFLICT   139    140       KL -> NV (IN REF. 1).
FT   CONFLICT   163    164       EL -> DV (IN REF. 1).
SQ   SEQUENCE   368 AA;  41654 MW;  6820F8DF CRC32;
```

```
  Query Match          17.7%;  Score 193;  DB 1;  Length 368;
  Best Local Similarity  34.5%;  Pred. No. 2.98e-18;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     61 AMCPFGCHCHLRVVQCSDLGLKSVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLYAL 119
          ::|| | |    |:|: ||:::| |: :| | | |:|| :  : :: |::| |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    120 VLVNNKISKIHEKAFSPLRKLQKLYIS 146
           | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    8
ID   PGS1_BOVIN     STANDARD;      PRT;   369 AA.
AC   P21809; P79259;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (LEUCINE-RICH PG I)
DE   (PG-S1).
GN   BGN.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AORTA;
RX   MEDLINE; 96113563.
RA   XU J.H., RADHAKRISHNAMURTHY B., SRINIVASAN S.R., BERENSON G.S.;
RT   "Primary structure of bovine aorta biglycan core protein deduced from
RT   cloned CDNA.";
RL   BIOCHEM. MOL. BIOL. INT. 37:263-272(1995).
RN   [2]
RP   SEQUENCE OF 38-369.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 89255324.
RA   NEAME P.J., CHOI H.U., ROSENBERG L.C.;
RT   "The primary structure of the core protein of the small, leucine-rich
RT   proteoglycan (PG I) from bovine articular cartilage.";
RL   J. BIOL. CHEM. 264:8653-8661(1989).
RN   [3]
RP   SEQUENCE OF 38-63.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 89123388.
RA   CHOI H.U., JOHNSON T.L., PAL S., TANG L.H., ROSENBERG L.C.,
RA   NEAME P.J.;
RT   "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT   DS-PGII, from bovine articular cartilage and skin isolated by octyl-
RT   sepharose chromatography.";
RL   J. BIOL. CHEM. 264:2876-2884(1989).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S82652; G1835865; -.
DR   PIR; A33701; A33701.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
```

```
KW   SIGNAL; REPEAT; LEUCINE-REPEAT.
FT   SIGNAL       1     19       POTENTIAL.
FT   PROPEP      20     37
FT   CHAIN       38    369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      93    316       LEUCINE-RICH REPEATS.
FT   REPEAT      93    106       LRR 1.
FT   REPEAT     117    130       LRR 2.
FT   REPEAT     141    154       LRR 3.
FT   REPEAT     162    175       LRR 4.
FT   REPEAT     186    199       LRR 5.
FT   REPEAT     211    224       LRR 6.
FT   REPEAT     232    245       LRR 7.
FT   REPEAT     256    269       LRR 8.
FT   REPEAT     280    288       LRR 9.
FT   REPEAT     303    316       LRR 10.
FT   CARBOHYD    42     42       GLYCOSAMINOGLYCAN.
FT   CARBOHYD    48     48       GLYCOSAMINOGLYCAN.
FT   CARBOHYD   271    271
FT   CARBOHYD   312    312
FT   DISULFID    64     77
FT   DISULFID   322    355
FT   CONFLICT   152    152       C -> V (IN REF. 2).
FT   CONFLICT   188    188       C -> E (IN REF. 2).
FT   CONFLICT   354    354       A -> R (IN REF. 2).
FT   CONFLICT   368    369       KK -> Y (IN REF. 2).
SQ   SEQUENCE   369 AA;  41509 MW;  F1CC673B CRC32;

  Query Match          17.7%;  Score 193;  DB 1;  Length 369;
  Best Local Similarity  34.5%;  Pred. No. 2.98e-18;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     62 AMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLYAL 120
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    121 VLVNNKISKIHEKAFSPLRKLQKLYIS 147
            | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT    9
ID   PGS2_CHICK     STANDARD;      PRT;   357 AA.
AC   P28675;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=WHITE LEGHORN; TISSUE=CORNEA;
RX   MEDLINE; 92296755.
RA   LI W., VERGNES J.P., CORNUET P.K., HASSEL J.R.;
RT   "cDNA clone to chick corneal chondroitin/dermatan sulfate
RT   proteoglycan reveals identity to decorin.";
RL   ARCH. BIOCHEM. BIOPHYS. 296:190-197(1992).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63797; G62888; -.
DR   PIR; S22197; S22197.
DR   PIR; S24317; S24317.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     16
FT   PROPEP      17     30
FT   CHAIN       31    357       BONE PROTEOGLYCAN II.
FT   DOMAIN      75    306       LEUCINE-RICH REPEATS.
FT   REPEAT      75     96       LRR 1.
FT   REPEAT      97    120       LRR 2.
FT   REPEAT     121    143       LRR 3.
FT   REPEAT     144    165       LRR 4.
FT   REPEAT     166    191       LRR 5.
FT   REPEAT     192    215       LRR 6.
FT   REPEAT     216    236       LRR 7.
FT   REPEAT     237    260       LRR 8.
FT   REPEAT     261    283       LRR 9.
FT   REPEAT     284    306       LRR 10.
FT   CARBOHYD    34     34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   209    209       POTENTIAL.
FT   CARBOHYD   260    260       POTENTIAL.
FT   DISULFID    52     65       BY SIMILARITY.
FT   DISULFID   311    344       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  39687 MW;  48F51E32 CRC32;

  Query Match          17.6%;  Score 192;  DB 1;  Length 357;
  Best Local Similarity  33.7%;  Pred. No. 4.66e-18;
  Matches   30;  Conservative   22;  Mismatches  36;  Indels   1;  Gaps   1;

Db     48 FGPVCPFRCQCHLRVVQCSDLGLERVPKDLPPDTTLLDLQNNKITEIK-EGDFKNLKNLH 106
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    107 ALILVNNKISKISPAAFAPLKKLERLYLS 135
              | | || |: |:  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   10
ID   PGS1_CANFA     STANDARD;      PRT;   369 AA.
AC   O02678;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
GN   BGN.
OS   CANIS FAMILIARIS (DOG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GLANT T.T.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC       CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC   -!- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC       EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U83140; G1916846; -.
DR   PFAM; PF00560; LRR; 5.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     19       POTENTIAL.
FT   PROPEP      20     37       BY SIMILARITY.
FT   CHAIN       38    369       BONE/CARTILAGE PROTEOGLYCAN I.
FT   DOMAIN      72    343       LEUCINE-RICH REPEATS.
FT   REPEAT      72     95       LRR 1.
FT   REPEAT      96    116       LRR 2.
FT   REPEAT     117    140       LRR 3.
FT   REPEAT     141    164       LRR 4.
FT   REPEAT     165    185       LRR 5.
FT   REPEAT     186    211       LRR 6.
FT   REPEAT     212    231       LRR 7.
FT   REPEAT     232    255       LRR 8.
FT   REPEAT     256    276       LRR 9.
FT   REPEAT     277    302       LRR 10.
FT   REPEAT     303    321       LRR 11.
FT   REPEAT     336    343       LRR 12.
FT   CARBOHYD    42     42       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD    48     48       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   271    271       POTENTIAL.
FT   CARBOHYD   312    312       POTENTIAL.
FT   DISULFID    64     77       BY SIMILARITY.
FT   DISULFID   322    355       BY SIMILARITY.
SQ   SEQUENCE   369 AA;  41566 MW;  F794CEEA CRC32;

  Query Match          17.3%;  Score 189;  DB 1;  Length 369;
  Best Local Similarity  35.2%;  Pred. No. 1.78e-17;
  Matches   31;  Conservative   21;  Mismatches  33;  Indels   3;  Gaps   3;

Db     62 AMCPFGCHCHLRVVQCSDLGLKAVPKEISP-DTMLLDLQNNDISELRAD-DFKGLHHLYA 119
          ::||  | |    |:|:  ||:::|  |:   | || |  |:|| :  : ::  | :|
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLL-LSGNNISTVDLNSNIHVLENLEX 126

Db    120 LVLVNNKISKIHEKAFSPLRKLQKLYIS 147
          | | || |: |::|:|  | ||: | ::
Qy    127 LDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   11
ID   PGS2_MOUSE     STANDARD;      PRT;   354 AA.
AC   P28654;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN) (PG40).
GN   DCN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS; TISSUE=FIBROBLAST;
RA   NAITOH Y., SUZUKI S.;
RL   SUBMITTED (JUL-1990) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95050610.
RA   SCHOLZEN T., SOLURSH M., SUZUKI S., REITER R., MORGAN J.L.,
RA   BUCHBERG A.M., SIRACUSA L.D., IOZZO R.V.;
RT   "The murine decorin. Complete cDNA cloning, genomic organization,
RT   chromosomal assignment, and expression during organogenesis and
RT   tissue differentiation.";
RL   J. BIOL. CHEM. 269:28270-28281(1994).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53929; G53669; -.
DR   PIR; S20812; S20812.
DR   MGD; MGI:94872; DCN.
DR   PFAM; PF00560; LRR; 5.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17     30
FT   CHAIN       31    354       BONE PROTEOGLYCAN II.
FT   DOMAIN      72    303       LEUCINE-RICH REPEATS.
FT   REPEAT      72     93       LRR 1.
FT   REPEAT      94    117       LRR 2.
FT   REPEAT     118    140       LRR 3.
FT   REPEAT     141    162       LRR 4.
FT   REPEAT     163    188       LRR 5.
FT   REPEAT     189    212       LRR 6.
FT   REPEAT     213    233       LRR 7.
FT   REPEAT     234    257       LRR 8.
FT   REPEAT     258    280       LRR 9.
FT   REPEAT     281    303       LRR 10.
FT   DISULFID    49     62       BY SIMILARITY.
FT   DISULFID   308    341       BY SIMILARITY.
FT   CARBOHYD    34     34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   206    206       POTENTIAL.
FT   CARBOHYD   241    241       POTENTIAL.
FT   CARBOHYD   257    257       POTENTIAL.
FT   CARBOHYD   298    298       POTENTIAL.
SQ   SEQUENCE   354 AA;  39809 MW;  F675E543 CRC32;

  Query Match          16.8%;  Score 184;  DB 1;  Length 354;
  Best Local Similarity  34.4%;  Pred. No. 1.63e-16;
  Matches   33;  Conservative   23;  Mismatches  37;  Indels   3;  Gaps   3;

Db     39 YDPDNPLI-SMCPYRCQCHLRVVQCSDLGLDKVPWDFPPDTTLLDLQNNKITEIK-EGAF 96
          | :|: :| |:|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :: :
Qy     60 Y-ADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNI 118

Db     97 KNLKDLHTLILVNNKISKISPEAFKPLVKLERLYLS 132
             | :|  | | || |: |:  :|  | ||  | |:
Qy    119 HVLENLEXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   12
ID   PGS2_RAT       STANDARD;      PRT;   354 AA.
AC   Q01129;
DT   01-APR-1993 (REL. 25, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN) (PG40) (DERMATAN
DE   SULFATE PROTEOGLYCAN-II) (DSPG).
GN   DCN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=UTERUS;
```

```
RX   MEDLINE; 93003331.
RA   ABRAMSON S.R., WOESSNER J.F.;
RT   "cDNA sequence for rat dermatan sulfate proteoglycan-II (decorin).";
RL   BIOCHIM. BIOPHYS. ACTA 1132:225-227(1992).
RN   [2]
RP   SEQUENCE OF 11-354 FROM N.A.
RX   MEDLINE; 93154359.
RA   ASUNDI V.K., DREHER K.L.;
RT   "Molecular characterization of vascular smooth muscle decorin:
RT   deduced core protein structure and regulation of gene expression.";
RL   EUR. J. CELL BIOL. 59:314-321(1992).
RN   [3]
RP   SEQUENCE OF 31-48 AND 171-191.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=UTERUS;
RX   MEDLINE; 89350825.
RA   KOKENYESI R., WOESSNER J.F.;
     "Purification and characterization of a small dermatan sulphate
     proteoglycan implicated in the dilatation of the rat uterine
     cervix.";
RL   BIOCHEM. J. 260:413-419(1989).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- FUNCTION: IMPLICATED IN THE DILATATION OF THE RAT CERVIX.
CC   -!- DEVELOPMENTAL STAGE: THE AMOUNT OF DSPG PER CERVIX INCREASES
CC       4-FOLD DURING PREGNANCY, THEN FALLS PRECIPITOUSLY WITHIN 1 DAY
CC       POST PATRUM.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z12298; G57550; -.
DR   EMBL; X59859; G56057; -.
DR   PIR; S29145; S29145.
DR   PFAM; PF00560; LRR; 5.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; CONNECTIVE TISSUE; EXTRACELLULAR MATRIX; PROTEOGLYCAN;
KW   REPEAT; LEUCINE-REPEAT; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30
FT   CHAIN       31   354       BONE PROTEOGLYCAN II.
FT   DOMAIN      72   303       LEUCINE-RICH REPEATS.
FT   REPEAT      72    93       LRR 1.
FT   REPEAT      94   117       LRR 2.
FT   REPEAT     118   140       LRR 3.
FT   REPEAT     141   162       LRR 4.
FT   REPEAT     163   188       LRR 5.
FT   REPEAT     189   212       LRR 6.
FT   REPEAT     213   233       LRR 7.
FT   REPEAT     234   257       LRR 8.
FT   REPEAT     258   280       LRR 9.
FT   REPEAT     281   303       LRR 10.
FT   DISULFID    49    62       BY SIMILARITY.
FT   DISULFID   308   341       BY SIMILARITY.
FT   CARBOHYD    34    34       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   184   184       GLYCOSAMINOGLYCAN (BY SIMILARITY).
FT   CARBOHYD   206   206       POTENTIAL.
FT   CARBOHYD   241   241       POTENTIAL.
FT   CARBOHYD   257   257       POTENTIAL.
FT   CARBOHYD   298   298       POTENTIAL.
SQ   SEQUENCE   354 AA;  39805 MW;  513FB9C1 CRC32;
```

```
  Query Match          16.8%;  Score 183;  DB 1;  Length 354;
  Best Local Similarity  34.4%;  Pred. No. 2.53e-16;
  Matches   33;  Conservative   23;  Mismatches  37;  Indels   3;  Gaps   3;

Db     39 YDPDNPLI-SMCPYRCQCHLRVVQCSDLGLDKVPWEFPPDTTLLDLQNNKITEIK-EGAF 96
          | :|: :| |:|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :: :
Qy     60 Y-ADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNI 118

Db     97 KNLKDLHTLILVNNKISKISPEAFKPLVKLERLYLS 132
            | :|  | | || |: |:  :|  | ||  | |:
Qy    119 HVLENLEXLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT  13
ID   GPBB_PAPCY     STANDARD;      PRT;   208 AA.
AC   Q04785;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN IB BETA CHAIN PRECURSOR (GP-IB BETA) (GPIBB)
DE   (CD42B-BETA) (CD42C).
OS   PAPIO CYNOCEPHALUS (YELLOW BABOON).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93273245.
RA   HAYZER D.J., SHOJI M., KIM T.M., RUNGE M.S., HANSON S.R.;
RT   "Characterization of a cDNA encoding the beta-chain of baboon
RT   receptor glycoprotein GPIb.";
RL   GENE 127:271-272(1993).
CC   -!- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,
CC       PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON
CC       WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM.
CC   -!- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS
CC       COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION OF THE
CC       MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET GLYCOPROTEIN IX
CC       (GP-IX) AND DISSOCIATION OF GP-IB FROM THE ACTIN-BINDING PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L05927; G176582; -.
DR   PFAM; PF00560; LRR; 1.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; HEMOSTASIS; BLOOD COAGULATION;
KW   SIGNAL; PHOSPHORYLATION; CELL ADHESION; LEUCINE-REPEAT.
FT   SIGNAL       1    26       BY SIMILARITY.
FT   CHAIN       27   208       PLATELET GLYCOPROTEIN IB BETA CHAIN.
FT   DOMAIN      27   147       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   148   172       BY SIMILARITY.
FT   DOMAIN     173   208       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    66    66       POTENTIAL.
FT   MOD_RES    193   193       PHOSPHORYLATION (BY CAPK)
FT                              (BY SIMILARITY).
SQ   SEQUENCE   208 AA;  21984 MW;  067DDEA5 CRC32;

  Query Match          14.8%;  Score 162;  DB 1;  Length 208;
  Best Local Similarity  38.5%;  Pred. No. 2.24e-12;
  Matches   25;  Conservative   18;  Mismatches  19;  Indels   3;  Gaps   2;

Db     26 CPAPCSCAGTLVDCGRRGLTWASLPTSFPVHTTELVLTGNNLTALP-SGLLDALPAVRTA 84
          ||: | | || ||| :|||   ::|||:|  :|:|:|:|||::::  :: :  |  :
Qy     70 CPTQCDCYGTTVDCNKRGLN--TIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127
```

```
Db     85 HLGAN 89
          |: |
Qy    128 DLSNN 132


RESULT   14
ID   GPBB_HUMAN     STANDARD;      PRT;   206 AA.
AC   P13224;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN IB BETA CHAIN PRECURSOR (GP-IB BETA) (GPIBB)
DE   (CD42B-BETA) (CD42C).
GN   GP1BB.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88176901.
RA   LOPEZ J.A., CHUNG D.W., FUJIKAWA K., HAGEN F.S., DAVIE E.W.,
RA   ROTH G.J.;
RT   "The alpha and beta chains of human platelet glycoprotein Ib are both
RT   transmembrane proteins containing a leucine-rich amino acid
RT   sequence.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 85:2135-2139(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 94292494.
RA   YAGI M., EDELHOFF S., DISTECHE C.M., ROTH G.J.;
RT   "Structural characterization and chromosomal location of the gene
RT   encoding human platelet glycoprotein Ib beta.";
RL   J. BIOL. CHEM. 269:17424-17427(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97174353.
RA   ZIEGER B., HASHIMOTO Y., WARE J.;
RT   "Alternative expression of platelet glycoprotein Ib(beta) mRNA from an
RT   adjacent 5' gene with an imperfect polyadenylation signal sequence.";
RL   J. CLIN. INVEST. 99:520-525(1997).
RN   [4]
RP   PHOSPHORYLATION, AND SEQUENCE OF 186-200.
RX   MEDLINE; 89359414.
RA   WARDELL M.R., REYNOLDS C.C., BERNDT M.C., WALLACE R.W., FOX J.E.B.;
RT   "Platelet glycoprotein Ib beta is phosphorylated on serine 166 by
RT   cyclic AMP-dependent protein kinase.";
RL   J. BIOL. CHEM. 264:15656-15661(1989).
RN   [5]
RP   SEQUENCE OF 27-40.
RX   MEDLINE; 87326405.
RA   CANFIELD V.A., OZOLS J., NUGENT D., ROTH G.J.;
RT   "Isolation and characterization of the alpha and beta chains of human
RT   platelet glycoprotein Ib.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 147:526-534(1987).
CC   -!- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,
CC       PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON
CC       WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM.
CC   -!- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS
CC       COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION OF THE
CC       MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET GLYCOPROTEIN IX
CC       (GP-IX) AND DISSOCIATION OF GP-IB FROM THE ACTIN-BINDING PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J03259; G306792; -.
DR   EMBL; AF006988; G2978512; -.
DR   EMBL; U59632; G1809266; -.
DR   PIR; B26864; B26864.
DR   PIR; A31929; A31929.
DR   MIM; 138720; -.
DR   PFAM; PF00560; LRR; 1.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; HEMOSTASIS; BLOOD COAGULATION;
KW   SIGNAL; PHOSPHORYLATION; CELL ADHESION; LEUCINE-REPEAT.
FT   SIGNAL       1    26
FT   CHAIN       27   206       PLATELET GLYCOPROTEIN IB BETA CHAIN.
FT   DOMAIN      27   147       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   148   172       POTENTIAL.
FT   DOMAIN     173   206       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      60    83       LRR.
FT   CARBOHYD    66    66       POTENTIAL.
FT   MOD_RES    191   191       PHOSPHORYLATION (BY CAPK).
SQ   SEQUENCE   206 AA;  21717 MW;  A9591D41 CRC32;

  Query Match          14.6%;  Score 159;  DB 1;  Length 206;
  Best Local Similarity  38.5%;  Pred. No. 7.95e-12;
  Matches   25;  Conservative   17;  Mismatches  20;  Indels   3;  Gaps   2;

Db     26 CPAPCSCAGTLVDCGRRGLTWASLPTAFPVDTTELVLTGNNLTALP-PGLLDALPALRTA 84
          ||: | | || ||| :|||   ::||::|  :|:|:|:|||::::   : :  |  |
Qy     70 CPTQCDCYGTTVDCNKRGLN--TIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db     85 HLGAN 89
          |: |
Qy    128 DLSNN 132


RESULT   15
ID   GPBB_MOUSE     STANDARD;      PRT;   206 AA.
AC   P56400;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN IB BETA CHAIN PRECURSOR (GP-IB BETA) (GPIBB).
GN   GP1BB.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97403789.
RA   KITAGUCHI T., MURATA M., ANBO H., MORIKI T., IKEDA Y.;
RT   "Characterization of the gene encoding mouse platelet glycoprotien Ib
RT   beta.";
RL   THROMB. RES. 87:235-244(1997).
CC   -!- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,
CC       PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON
CC       WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS
CC       COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION OF THE
CC       MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET GLYCOPROTEIN IX
CC       (GP-IX) AND DISSOCIATION OF GP-IB FROM THE ACTIN-BINDING PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

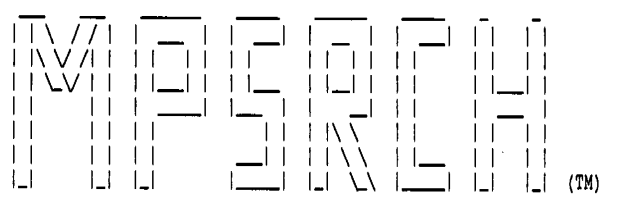
```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB001419; D1023289; -.
DR   MGD; MGI:107852; GP1BB.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; HEMOSTASIS; BLOOD COAGULATION;
KW   SIGNAL; PHOSPHORYLATION; CELL ADHESION; LEUCINE-REPEAT.
FT   SIGNAL       1    26       BY SIMILARITY.
FT   CHAIN       27   206       PLATELET GLYCOPROTEIN IB BETA CHAIN.
FT   DOMAIN      27   147       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   148   172       POTENTIAL.
FT   DOMAIN     173   206       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      60    83       LRR.
FT   CARBOHYD    66    66       POTENTIAL.
FT   MOD_RES    191   191       PHOSPHORYLATION (BY CAPK) (BY
FT                              SIMILARITY).
SQ   SEQUENCE   206 AA;  21762 MW;  4D3362F4 CRC32;

 Query Match           14.6%;  Score 159;  DB 1;  Length 206;
 Best Local Similarity  36.9%;  Pred. No. 7.95e-12;
 Matches   24;  Conservative   18;  Mismatches  20;  Indels   3;  Gaps   2;

Db     26 CPAPCSCAGTLVDCGRRGLTWASLPAAFPPDTTELVLTGNNLTALP-PGLLDALPALRAA 84
          ||: | | || ||| :|||   ::|:::|  :|:|:|:|||::::   : :  |  |
Qy     70 CPTQCDCYGTTVDCNKRGLN--TIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db     85 HLGAN 89
             |: |
Qy    128 DLSNN 132


Search completed: Fri May 28 09:05:09 1999
Job time : 27 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

ssrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:05:28 1999; MasPar time 12.78 Seconds
662.015 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-8
Description: (1-155) from US09191647.pep
Perfect Score: 1092
Sequence: 1 RNPXICDCNLQWLAQINLQK.....FINDKSFEKLSKLRELXLND 155

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 42.369; Variance 75.870; scale 0.558

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	497	45.5	1523	11	O88280	MEGF5.	1.49e-82
2	395	36.2	1531	11	O88279	MEGF4.	4.57e-60
3	213	19.5	167	5	Q21042	F59F3.7.	5.83e-22
4	202	18.5	224	5	O44086	ZK994.4 PROTEIN.	8.27e-20
5	201	18.4	420	4	Q13641	5T4 ONCOFETAL ANTIGEN	1.29e-19
6	197	18.0	811	4	O75139	KIAA0644 PROTEIN.	7.69e-19
7	195	17.9	360	6	O46542	DERMATAN SULFATE PROTE	1.87e-18
8	195	17.9	360	6	Q28888	DECORIN.	1.87e-18
9	193	17.7	369	6	O46390	BIGLYCAN PRECURSOR.	4.53e-18
10	192	17.6	372	6	O46403	BIGLYCAN.	7.05e-18
11	191	17.5	321	6	P79119	EPIPHYCAN.	1.10e-17
12	191	17.5	322	4	Q99645	DERMATAN SULFATE PROTE	1.10e-17
13	184	16.8	322	11	P70186	PROTEOGLYCAN PRECURSOR	2.37e-16
14	182	16.7	316	13	Q90944	PROTEOGLYCAN CORE PROT	5.67e-16
15	182	16.7	1385	5	Q26388	TLR=TOLL-LIKE RECEPTOR	5.67e-16
16	182	16.7	1389	5	Q24591	WHEELER.	5.67e-16
17	181	16.6	718	13	O73675	NEURONAL LEUCINE-RICH	8.76e-16
18	178	16.3	522	4	O43354	BAC CLONE GS099H08, CO	3.22e-15
19	173	15.8	1496	4	Q92626	MYELOBLAST KIAA0230 (F	2.78e-14
20	169	15.5	705	4	O43377	BAC CLONE RG118D07 FRO	1.54e-13
21	168	15.4	733	5	Q24250	TARTAN PROTEIN PRECURS	2.35e-13
22	167	15.3	516	4	O43300	KIAA0416.	3.60e-13
23	167	15.3	707	11	P97860	LEUCINE-RICH REPEAT PR	3.60e-13
24	166	15.2	713	4	O75325	GLIOMA AMPLIFIED ON CH	5.50e-13
25	164	15.0	96	11	Q63156	DECORIN (FRAGMENT).	1.28e-12
26	163	14.9	423	11	O35103	OSTEOMODULIN.	1.95e-12
27	161	14.7	421	4	Q99983	OSTEOMODULIN, COMPLETE	4.53e-12
28	161	14.7	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	4.53e-12
29	159	14.6	411	4	Q14422	GLYCOPROTEIN IB BETA.	1.05e-11
30	159	14.6	907	4	O75473	ORPHAN G PROTEIN-COUPL	1.05e-11
31	157	14.4	716	11	Q61809	LEUCINE-RICH-REPEAT PR	2.41e-11
32	156	14.3	1535	5	Q23991	PEROXIDASIN PRECURSOR.	3.65e-11
33	151	13.8	892	5	P91644	KEK2 PRECURSOR (FRAGME	2.88e-10
34	150	13.7	739	4	O75094	MEGF5 (FRAGMENT).	4.34e-10
35	147	13.5	428	4	O14498	ISLR PRECURSOR.	1.48e-09
36	144	13.2	661	11	Q62192	LYMPHOCYTE ANTIGEN 78	4.99e-09
37	143	13.1	683	5	Q22187	T05A1.3 PROTEIN.	7.46e-09
38	142	13.0	526	10	O22753	PREDICTED LEUCINE-RICH	1.12e-08
39	142	13.0	880	5	P91643	KEK1 PRECURSOR.	1.12e-08
40	139	12.7	135	6	O46377	BIGLYCAN (FRAGMENT).	3.70e-08
41	139	12.7	230	6	Q29396	LEUCINE-RICH GLYCOPROT	3.70e-08
42	138	12.6	4293	11	O08852	POLYCYSTIC KIDNEY DISE	5.51e-08
43	136	12.5	358	11	O55226	CHONDROADHERIN.	1.22e-07
44	136	12.5	422	6	O77742	OSTEOADHERIN.	1.22e-07
45	136	12.5	661	4	Q99467	LYMPHOCYTE ANTIGEN 64	1.22e-07

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 45.5%; Score 497; DB 11; Length 1523;
Best Local Similarity 40.9%; Pred. No. 1.49e-82;
Matches 63; Conservative 45; Mismatches 44; Indels 2; Gaps 2;

```
Db   436 QNPFVCDCHLKWLAD-YLQDNPIETSGARCSSPRRLANKRISQIKSKKFRCSGSEDYRNR 494
            :|| :|||:| |||:  || | |||||||  |:|| :|::: : ::||:| ||| : :
Qy     1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSESFVSM 59

Db   495 FSSECFMDLVCPEKCRCEGTIVDCSNQKLSRIPSHLPEYTTDLRLNDNDIAVLEATGIFK 554
            ::  ||:| :||  | | || |||::: |: ||: :| ::|:| |: |:|: ::  : :
Qy    60 YADSCFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIH 119

Db   555 KLPNLRKINLSNNRIKEVREGAFDGAAGVQELML 588
           | ||  ::||||:|  : : :|:  : ::|| |
Qy   120 VLENLEXLDLSNNHITFINDKSFEKLSKLRELXL 153
```

RESULT 2

```
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match          36.2%;  Score 395;  DB 11;  Length 1531;
  Best Local Similarity  39.4%;  Pred. No. 4.57e-60;
  Matches   63;  Conservative   41;  Mismatches  48;  Indels   8;  Gaps   7;

Db   438 QNPFICDCNLKWLADF-LRTNPIETTGARCASPRRLANKRIGQIKSKKFRCSAKEQYFIP 496
         :|| |||||| |||:: |: | |||:|||| |:|| :|::: : ::||:| : |  |::
Qy     1 RNPXICDCNLQWLAQINLQKN-IETSGARCEQPKRLRKKKFATLPPNKFKCKGSES-FVS 58

Db   497 GTEDYHLNSECTSDVACPHKCRCEASVVECSGLKLSKIPERIPQSTTELRLNNNEISILE 556
               |  :| |  |  ||  | | :: |:|:   |: ||  ||: :|:| |: |:|| ::
Qy    59 -M--YA-DS-CFIDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVD 113

Db   557 ATGLFKKLSHLKKINLSNNKVSEIEDGTFEGATSVSELHL 596
           : :  | :|  ::|||| :: |:| :||  : : || |
Qy   114 LNSNIHVLENLEXLDLSNNHITFINDKSFEKLSKLRELXL 153


RESULT   3
ID   Q21042      PRELIMINARY;      PRT;  167 AA.
AC   Q21042;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F59F3.7.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KERSHAW J.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z68005; G1070069; -.
DR   PFAM; PF00560; LRR; 2.
SQ   SEQUENCE   167 AA;  18567 MW;  4F0F09BF CRC32;

  Query Match          19.5%;  Score 213;  DB 5;  Length 167;
  Best Local Similarity  36.0%;  Pred. No. 5.83e-22;
  Matches   32;  Conservative   25;  Mismatches  31;  Indels   1;  Gaps   1;

Db    13 LSTACPSECRCAGLDVHCEGKNLTAIPGHIPIATTNLYFSNNLLNSLS-KSNFQALPNLQ 71
           : : ||::| | |  | |: : | :||  ||  :| | :| | :::: :||:: | ||:
Qy    66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    72 YLDLSNNSIRDIEETLLDSFPGLKYLDLS 100
            |||||| |  |::  :: :: |: | |:
Qy   126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   4
ID   O44086      PRELIMINARY;      PRT;  224 AA.
AC   O44086;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZK994.4 PROTEIN.
GN   ZK994.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., WOHLDMANN P.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF022977; G2668616; -.
SQ   SEQUENCE   224 AA;  25885 MW;  3720F7F5 CRC32;

  Query Match          18.5%;  Score 202;  DB 5;  Length 224;
  Best Local Similarity  32.6%;  Pred. No. 8.27e-20;
  Matches   30;  Conservative   24;  Mismatches  34;  Indels   4;  Gaps   3;

Db    17 FVAGLECPVECTCDKKGLVVDCSSSGLTRIPKNISRNVRSLVIRNNRIHKLK-RSDLEGF 75
           |: :: || :| |   |  |||:  ||  || :|:|    |::  | |  :  |::  :
Qy    65 FIDSI-CPTQCDCY--GTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVL 121

Db    76 NQLETLVLTHNKIKIIEENVLDHLPELKRLSL 107
           : || | |::| | :|::: :: |: |: | |
Qy   122 ENLEXLDLSNNHITFINDKSFEKLSKLRELXL 153


RESULT   5
ID   Q13641      PRELIMINARY;      PRT;  420 AA.
AC   Q13641;
```

```
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   5T4 ONCOFETAL ANTIGEN PRECURSOR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 94179356.
RA   MYERS K.A., RAHI-SAUND V., DAVISON M.D., YOUNG J.A., CHEATER A.J.,
RA   STERN P.L.;
RT   "Isolation of a cDNA encoding 5T4 oncofetal trophoblast glycoprotein.
RT   An antigen associated with metastasis contains leucine-rich
RT   repeats.";
RL   J. BIOL. CHEM. 269:9319-9324(1994).
DR   EMBL; Z29083; G435655; -.
DR   PFAM; PF00560; LRR; 4.
KW   SIGNAL.
FT   SIGNAL       1     31       POTENTIAL.
FT   CHAIN       32    420       5T4 ONCOFETAL ANTIGEN.
SQ   SEQUENCE   420 AA;  46031 MW;  43633112 CRC32;

  Query Match          18.4%;  Score 201;  DB 4;  Length 420;
  Best Local Similarity  31.5%;  Pred. No. 1.29e-19;
  Matches   28;  Conservative   25;  Mismatches  32;  Indels   4;  Gaps   4;

Db     62 CPALCECSEAARTVKCVNRNLTEVPTDLPAYVRNLFLTGNQLAVLPAGAFARRPPLAELA 121
          ||: |:|  :: || | :| |  :|| :| :   |:|:|| :: :   :   :  | :|
Qy     70 CPTQCDCY-GT-TVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNS-NIHV-LENLE 125

Db    122 ALNLSGSRLDEVRAGAFEHLPSLRQLDLS 150
           |:|| :::  :   :|| |: ||:| |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   6
ID   O75139      PRELIMINARY;      PRT;   811 AA.
AC   O75139;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KIAA0644 PROTEIN.
GN   KIAA0644.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98403880.
RA   ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA   NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA RES. 5:169-176(1998).
DR   EMBL; AB014544; D1032580; -.
SQ   SEQUENCE   811 AA;  88695 MW;  C8B8C147 CRC32;

  Query Match          18.0%;  Score 197;  DB 4;  Length 811;
  Best Local Similarity  38.3%;  Pred. No. 7.69e-19;
  Matches   36;  Conservative   19;  Mismatches  33;  Indels   6;  Gaps   5;

Db     24 EPVCPERCDCQHPQHLLCTNRGLRVVPKTSSLPSPHDVLTYSLGGNFITNIT-AFDFHRL 82
          :::|| :|||     : | :|||  :| || :|     |   |:|| |: :    ::| |
Qy     67 DSICPTQCDC-YGTTVDCNKRGLNTIP-TS-IPRFATQLL--LSGNNISTVDLNSNIHVL 121

Db     83 GQLRRLDLQYNQIRSLHPKTFEKLSRLEELYLGN 116
             |  |||  |:|  :: |:|||||:| || | :
Qy    122 ENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155


RESULT   7
ID   O46542      PRELIMINARY;      PRT;   360 AA.
AC   O46542;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   DERMATAN SULFATE PROTEOGLYCAN II.
OS   EQUUS CABALLUS (HORSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PERISSODACTYLA; EQUIDAE; EQUUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RICHARDSON D.W., DODGE G.R.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF038127; G2723947; -.
SQ   SEQUENCE   360 AA;  39939 MW;  A3C42C76 CRC32;

  Query Match          17.9%;  Score 195;  DB 6;  Length 360;
  Best Local Similarity  33.7%;  Pred. No. 1.87e-18;
  Matches   30;  Conservative   22;  Mismatches  36;  Indels   1;  Gaps   1;

Db     51 LGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNLH 109
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    110 ALILVNNKISKISPGAFTPLVKLERLYLS 138
           | | || |: |:  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   8
ID   Q28888      PRELIMINARY;      PRT;   360 AA.
AC   Q28888; Q28608;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DECORIN.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95122319.
RA   ZHAN Q., BURROWS R., CINTRON C.;
RT   "Cloning and in situ hybridization of rabbit decorin in corneal
RT   tissues.";
RL   INVEST. OPHTHALMOL. VIS. SCI. 36:206-215(1995).
RN   [2]
RP   SEQUENCE OF 38-358 FROM N.A.
RC   TISSUE=CARTILAGE;
RA   HERING T.M., KOLLAR J.;
RL   SUBMITTED (NOV-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; S76584; G913375; -.
DR   EMBL; U03394; G415647; -.
DR   PFAM; PF00560; LRR; 5.
SQ   SEQUENCE   360 AA;  39896 MW;  1FAB2F8E CRC32;

  Query Match          17.9%;  Score 195;  DB 6;  Length 360;
  Best Local Similarity  33.7%;  Pred. No. 1.87e-18;
  Matches   30;  Conservative   22;  Mismatches  36;  Indels   1;  Gaps   1;

Db     51 LGPVCPFRCQCHLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIK-DGDFKNLKNLH 109
          : ::|| :|:|    |:|:  ||: :|  :|  :| | |  |:|: :  :::  | ||
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    110 ALILVNNKISKISPGAFTPLVKLERLYLS 138
           | | || |: |:  :|  | ||  | |:
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   9
ID   O46390      PRELIMINARY;      PRT;   369 AA.
```

```
AC   O46390;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   BIGLYCAN PRECURSOR.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; CAPRINAE; OVIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CHOROID PLEXUS;
RA   BRUETT L., CLEMENTS J.E.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF034842; G2655356; -.
KW   SIGNAL.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       38    369       BIGLYCAN.
SQ   SEQUENCE   369 AA;  41523 MW;  A0A9F549 CRC32;

  Query Match           17.7%;  Score 193;  DB 6;  Length 369;
  Best Local Similarity  34.5%;  Pred. No. 4.53e-18;
  Matches   30;  Conservative   23;  Mismatches  33;  Indels   1;  Gaps   1;

Db     62 AMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNDISELR-KDDFKGLQHLYAL 120
          ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|  |
Qy     68 SICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXL 127

Db    121 VLVNNKISKIHEKAFSPLRKLQKLYIS 147
           | || |: |::|:|  | ||: | ::
Qy    128 DLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   10
ID   O46403      PRELIMINARY;      PRT;   372 AA.
AC   O46403;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   BIGLYCAN.
OS   EQUUS CABALLUS (HORSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PERISSODACTYLA; EQUIDAE; EQUUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RICHARDSON D.W., DODGE G.R.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF035934; G2662531; -.
SQ   SEQUENCE   372 AA;  41924 MW;  097E7BA6 CRC32;

  Query Match           17.6%;  Score 192;  DB 6;  Length 372;
  Best Local Similarity  33.7%;  Pred. No. 7.05e-18;
  Matches   30;  Conservative   24;  Mismatches  34;  Indels   1;  Gaps   1;

Db     63 FSAMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNEISELR-KDDFKGLQHLY 121
          : ::||  | |    |:|:  ||:::|  |:  :| | |  |:|| :  : ::  |::|
Qy     66 IDSICPTQCDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLE 125

Db    122 ALVLVNNKISKIHEKAFSPLRKLQKLYIS 150
           | | || |: |::|:|  | ||: | ::
Qy    126 XLDLSNNHITFINDKSFEKLSKLRELXLN 154


RESULT   11
ID   P79119      PRELIMINARY;      PRT;   321 AA.
AC   P79119;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   EPIPHYCAN.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97373567.
RA   JOHNSON H.J., ROSENBERG L., CHOI H.U., GARZA S., HOOK M., NEAME P.J.;
RT   "Characterization of epiphycan, a small proteoglycan with a
RT   leucine-rich repeat core protein.";
RL   J. BIOL. CHEM. 272:18709-18717(1997).
DR   EMBL; U77127; G1679745; -.
DR   PFAM; PF00560; LRR; 3.
SQ   SEQUENCE   321 AA;  36687 MW;  D1B0F07E CRC32;

  Query Match           17.5%;  Score 191;  DB 6;  Length 321;
  Best Local Similarity  34.1%;  Pred. No. 1.10e-17;
  Matches   28;  Conservative   27;  Mismatches  25;  Indels   2;  Gaps   2;

Db    120 CTCISTTVYCDDHELDAIP-PLPKNTAYFYSRFNRIKKIN-KNDFASLNDLRRIDLTSNL 177
          | | :||| |: : |::|| ::|: :: :    | |  :: :::: |::|  :||::|
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXLDLSNNH 133

Db    178 ISEIDEDAFRKLPQLRELVLRD 199
           |: |:: :| ||: |||| | |
Qy    134 ITFINDKSFEKLSKLRELXLND 155


RESULT   12
ID   Q99645      PRELIMINARY;      PRT;   322 AA.
AC   Q99645;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DERMATAN SULFATE PROTEOGLYCAN 3.
GN   DSPG3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CHONDROCYTE;
RX   MEDLINE; 97131519.
RA   DEERE M., JOHNSON J., GARZA S., HARRISON W.R., YOON S.J.,
RA   ELDER F.F.B., KUCHERLAPATI R., HOOK M., HECHT J.T.;
RT   "Characterization of human DSPG3, a small dermatan sulfate
RT   proteoglycan.";
RL   GENOMICS 38:399-404(1996).
DR   EMBL; U59111; G1794209; -.
DR   PFAM; PF00560; LRR; 3.
SQ   SEQUENCE   322 AA;  36607 MW;  DCB533EB CRC32;

  Query Match           17.5%;  Score 191;  DB 4;  Length 322;
  Best Local Similarity  34.1%;  Pred. No. 1.10e-17;
  Matches   28;  Conservative   26;  Mismatches  26;  Indels   2;  Gaps   2;

Db    121 CTCISTTVYCDDHELDAIP-PLPKNTAYFYSRFNRIKKIN-KNDFASLSDLKRIDLTSNL 178
          | | :||| |: : |::|| ::|: :: :    | |  :: :::: | :|  :||::|
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXLDLSNNH 133

Db    179 ISEIDEDAFRKLPQLRELVLRD 200
           |: |:: :| ||: |||| | |
Qy    134 ITFINDKSFEKLSKLRELXLND 155


RESULT   13
ID   P70186      PRELIMINARY;      PRT;   322 AA.
AC   P70186;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   PROTEOGLYCAN PRECURSOR.
GN   PG-LB.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=BALB/C; TISSUE=EPIPHYSEAL CARTILAGE;
RX   MEDLINE; 96433109.
RA   KURITA K., SHINOMURA T., UJITA M., ZAKO M., KIDA D., IWATA H.,
RA   KIMATA K.;
RT   "Occurence of PG-Lb, a leucine-rich small chondroitin/dermatan
RT   sulphate proteoglycan in mammalian epiphyseal cartilage: molecular
RT   cloning and sequence analysis of the mouse cDNA.";
RL   BIOCHEM. J. 318:909-914(1996).
DR   EMBL; D78274; D1011999; -.
DR   MGD; MGI:107942; DSPG3.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL; PROTEOGLYCAN.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26    322       PROTEOGLYCAN.
SQ   SEQUENCE   322 AA;  36762 MW;  AEA1D1E8 CRC32;

  Query Match          16.8%;  Score 184;  DB 11;  Length 322;
  Best Local Similarity  34.1%;  Pred. No. 2.37e-16;
  Matches   28;  Conservative   27;  Mismatches  25;  Indels   2;  Gaps   2;

Db    121 CTCISTTVYCDDHELDAIP-PLPKKTTYFYSRFNRIKKIN-KNDFASLNDLKRIDLTSNL 178
          | | :||| |: : |::|| ::|: :| :    | |  :: :::: |::|  :||::|
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSGNNISTVDLNSNIHVLENLEXLDLSNNH 133

Db    179 ISEIDEDAFRKLPHLQELVLRD 200
          |: |:: :| ||: |:|| | |
Qy    134 ITFINDKSFEKLSKLRELXLND 155


RESULT   14
ID   Q90944      PRELIMINARY;      PRT;   316 AA.
AC   Q90944;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   PROTEOGLYCAN CORE PROTEIN PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=TIBIAL CARTILAGE;
RX   MEDLINE; 92112771.
RA   SHINOMURA T., KIMATA K.;
RT   "Proteoglycan-Lb, a small dermatan sulfate proteoglycan expressed in
RT   embryonic chick epiphyseal cartilage, is structurally related to
RT   osteoinductive factor.";
RL   J. BIOL. CHEM. 267:1265-1270(1992).
DR   EMBL; D10485; D1001838; -.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    316       PROTEOGLYCAN CORE PROTEIN.
SQ   SEQUENCE   316 AA;  35856 MW;  CEAC004B CRC32;

  Query Match          16.7%;  Score 182;  DB 13;  Length 316;
  Best Local Similarity  36.1%;  Pred. No. 5.67e-16;
  Matches   30;  Conservative   23;  Mismatches  26;  Indels   4;  Gaps   4;

Db    115 CTCLGTTVYCDDRELDAVP-PLPK-NTMYFYSRYNRIRKIN-KNDFANLNNLKRIDLTAN 171
          | | |||| |: | |:::| ::|:  |  : |  | |  :: :::: |:|| :||: |
Qy     74 CDCYGTTVDCNKRGLNTIPTSIPRFATQLLLSG-NNISTVDLNSNIHVLENLEXLDLSNN 132

Db    172 LISEIHEDAFRRLPQLLELVLRD 194
          |: |:: :| :|: | || | |
Qy    133 HITFINDKSFEKLSKLRELXLND 155


RESULT   15
ID   Q26388      PRELIMINARY;      PRT;  1385 AA.
AC   Q26388;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TLR=TOLL-LIKE RECEPTOR.
GN   TLR.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95151581.
RA   CHIANG C., BEACHY P.A.;
RT   "Expression of a novel Toll-like gene spans the parasegment boundary
RT   and contributes to hedgehog function in the adult eye of
RT   Drosophila.";
RL   MECH. DEV. 47:225-239(1994).
DR   EMBL; S76155; G913248; -.
DR   FLYBASE; FBgn0004364; 18w.
DR   PFAM; PF00560; LRR; 13.
SQ   SEQUENCE  1385 AA;  154848 MW;  60273533 CRC32;

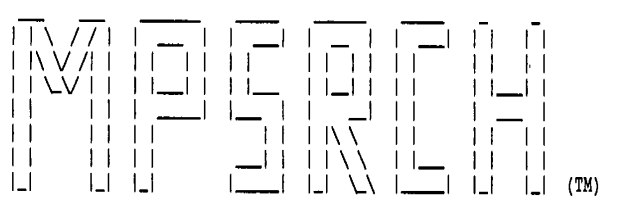
  Query Match          16.7%;  Score 182;  DB 5;  Length 1385;
  Best Local Similarity  29.1%;  Pred. No. 5.67e-16;
  Matches   50;  Conservative   44;  Mismatches  59;  Indels  19;  Gaps  14;

Db    679 NPFECDCSMEWLQRINNLTTRQHPHVVDLGNIECLMPHSRSAPLRPLASLSASDFVCKYE 738
          ||  |||:::|| :|| |   |  :: : :   |  | |    : :|:|::: | ||
Qy      2 NPXICDCNLQWLAQIN-L---QK-NI-ETSGARCEQPK-R-LRKKKFATLPPNKFKCK-G 52

Db    739 SHCPPTCHCCEYEQCECEVICPGNCSCFHDATWATNIVDCGRQDLAALPNRIPQDVSDLY 798
          |  :       : |  : |||  | |:  :|  |  ||| :: | ::|  ||:  ::|
Qy     53 SE-SFVSMY-A-DSCFIDSICPTQCDCY--GT--T--VDCNKRGLNTIPTSIPRFATQLL 103

Db    799 LDGNNMPELEV-GHLTGRRNLRALYLNASNLMTLQNGSLAQLVNLRVLHLEN 849
          | |||:: ::: :::    ||  | |: :::  : : |:  | :|| | |::
Qy    104 LSGNNISTVDLNSNIHVLENLEXLDLSNNHITFINDKSFEKLSKLRELXLND 155


Search completed: Fri May 28 09:06:08 1999
Job time : 40 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:08:00 1999; MasPar time 27.18 Seconds
575.102 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-9
Description: (1-735) from US09191647.pep
Perfect Score: 5438
Sequence: 1 SNKNLTSFPSRIPFDTTELY.....TVHIIRQCQCEPTKSVLSEK 735

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 36.837; Variance 203.463; scale 0.181

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1813	33.3	1480	5	R25079	Drosophila SLIT prote	2.28e-126
2	1673	30.8	1534	30	W46966	Amino acid sequence o	1.47e-115
3	533	9.8	727	21	W11719	C-Delta-1 polypeptide	1.02e-28
4	533	9.8	740	21	W00876	C-Delta-1 polypeptide	1.02e-28
5	525	9.7	520	25	W18348	Proliferation and dif	3.98e-28
6	527	9.7	685	37	W80813	Nucleotide sequence o	2.83e-28
7	525	9.7	702	25	W18349	Proliferation and dif	3.98e-28
8	527	9.7	722	21	W11720	M-Delta-1 polypeptide	2.83e-28
9	525	9.7	723	25	W18353	Proliferation and dif	3.98e-28
10	521	9.6	833	6	R28960	Delta D11.	7.87e-28
11	489	9.0	1036	25	W18351	Proliferation and dif	1.81e-25
12	489	9.0	1187	25	W18352	Proliferation and dif	1.81e-25
13	489	9.0	1218	29	W44301	Human serrate 1.	1.81e-25
14	489	9.0	1218	19	W05833	Human Serrate-1 (HJ1)	1.81e-25
15	484	8.9	1208	28	W40827	Human Jagged protein.	4.22e-25
16	482	8.9	1872	36	W68510	Partial human Notch-3	5.93e-25
17	486	8.9	2321	36	W49698	Human Notch3 protein.	3.01e-25
18	471	8.7	1193	19	W05835	Chick Serrate.	3.82e-24
19	475	8.7	1218	25	W18354	Proliferation and dif	1.94e-24
20	455	8.4	1055	29	W44298	Human serrate 2 prote	5.70e-23
21	455	8.4	1212	29	W44299	Human serrate 2.	5.70e-23
22	455	8.4	1257	19	W05834	Human Serrate-2 (HJ2)	5.70e-23
23	445	8.2	612	28	W39256	Human partial mature	3.08e-22
24	445	8.2	737	28	W39257	Human membrane protei	3.08e-22
25	420	7.7	660	21	W11725	H-Delta-1 polypeptide	2.06e-20
26	416	7.6	383	10	R56166	Neuroendocrine tumor	4.02e-20
27	401	7.4	385	10	R56167	Neuroendocrine tumor	4.95e-19
28	398	7.3	1404	7	R38304	Sequence of a serrate	8.18e-19
29	389	7.2	157	21	W11730	H-Delta-1 polypeptide	3.67e-18
30	374	6.9	473	17	R86869	Adhesive protein.	4.46e-17
31	353	6.5	196	5	R29102	Drosophila SLIT prote	1.45e-15
32	338	6.2	2189	1	R05222	Antigen GX5401FL enco	1.72e-14
33	331	6.1	228	30	W46967	Amino acid sequence o	5.42e-14
34	311	5.7	2707	24	W27161	Mouse receptor ME2.	1.43e-12
35	307	5.6	2409	3	R12609	Versican.	2.75e-12
36	300	5.5	77	6	R28962	ELR-11 and -12.	8.59e-12
37	286	5.3	1091	27	W41641	Sequence used in dete	8.31e-11
38	275	5.1	751	10	R53088	Human masking protein	4.90e-10
39	275	5.1	752	10	R53087	Human masking protein	4.90e-10
40	275	5.1	756	10	R53086	Human masking protein	4.90e-10
41	275	5.1	845	10	R53089	Human masking protein	4.90e-10
42	273	5.0	1257	9	R46627	Neurocan core protein	6.76e-10
43	264	4.9	1355	3	R14584	TGF beta 1 binding pr	2.86e-09
44	267	4.9	1712	4	R22461	Masking protein high	1.77e-09
45	261	4.8	240	33	W64219	Human secreted protei	4.62e-09

ALIGNMENTS

RESULT 1
ID R25079 standard; Protein; 1480 AA.
AC R25079;
DT 05-JAN-1993 (first entry)
DE Drosophila SLIT protein involved in axon pathway development.
KW Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW midline glial cells; axonogenesis; cell-cell interaction; ss.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT peptide 1..36
FT /label= signal
FT domain 73..294
FT /label= Flank_LRR_Flank_1
FT /note= "mediates adhesive events"
FT domain 295..518
FT /label= Flank-LRR-Flank_2
FT /note= "mediates adhesive events"
FT domain 519..714
FT /label= Flank_LRR_Flank_3
FT /note= "mediates adhesive events"
FT domain 715..910
FT /label= Flank_LRR_Flank_4
FT /note= "mediates adhesive events"
FT region 911..1150
FT /label= Tandem_EGF_like_repeats
FT /note= "involved in protein-protein interactions"
FT region 1353..1393
FT /label= 7th_EGF_like_repeat
FT /note= "involved in receptor-ligand interactions"
FT region 1394..1404
FT /label= alternative_splice_segment
FT /note= "developmentally regulated"
FT region 1405..1480
FT /label= C-terminal_region
PN WO9210518-A.
PD 25-JUN-1992.
PF 27-NOV-1991; U09055.
PR 07-DEC-1990; US-624135.
PA (UYYA) UNIV YALE.

PI Artavanis-Tsakonas S, Rothberg JM;
DR WPI; 92-234590/28.
DR N-PSDB; Q25811.
PT SLIT protein and sequence elements for treating
PT neuro-degenerative disease - useful for Alzheimer's disease,
PT nerve damage and Parkinson's disease, for diagnosis of cancer
PS Claim 1; Page 84-89; 122pp; English.
CC The SLIT protein is necessary for normal development of the midline
CC of the CNS, partic. the midline glial cells, and for the
CC concomitant formation of the commisural axon pathways. The process
CC is dependent on the level of SLIT protein expression. It appears
CC that SLIT protein is excreted by the midline glial cells where it
CC is synthesised and is eventually associated with the surfaces of
CC axons that traverse tham. The SLIT protein is tightly localised to
CC the muscle attachment sites and to the sites of contact between
CC adjacent pairs of cardioblasts as they coalesce to form the lumen
CC of the larval heart. The SLIT protein defines a new set of
CC molecules (TAGONS) which play a key role in axon outgrowth and
CC pathfinding. SLIT can be used as a nerve regenerative in
CC neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC region-Flank" domains, the C-terminal region and part of the
CC alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC and at least 1 EGF-like repeat element from the SLIT protein.
CC See also R29102.
SQ Sequence 1480 AA;

Query Match 33.3%; Score 1813; DB 5; Length 1480;
Best Local Similarity 41.1%; Pred. No. 2.28e-126;
Matches 253; Conservative 145; Mismatches 196; Indels 21; Gaps 15;

Db 729 srnqlkeiprgipaetsqvylesneieqihyerirhlrsltrldlsnnqitilsnytfan 788
Qy 1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRLYSLTKLDLSHNRLISLENNTFSN 60

Db 789 ltklsrliisynklqclqrhalsglnnlrvvslhgnrismlpeasfedlkslthialgsn 848
Qy 61 LTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGSN 120

Db 849 plycdcglkwfsdwikldyvepgiarcaepeqmkdklilstpsssfvcrgrvrndilakc 908
Qy 121 SLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKC 180

Db 909 nacfeqpcqnqaqcvalpqreyqclcqpgyhgkhcefmidacygnpcrnnatctvleegr 968
Qy 181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db 969 frcgcapgytgarcetniddcl-g--ei--kcqnnat-c---i-d--g--vesykcecqp 1014
Qy 241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db 1015 gfsgefcdtkiqfcspefnpcangakcmdhfthyscdcqagfhgtnctdniddcqnhmcq 1074
Qy 301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db 1075 nggtcvdgindyqcrcpddytgkyceghnmismmypqtspcqnheckhgvcfqpnaqgsd 1134
Qy 361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECV-ASQNSSD 419

Db 1135 ylcrchpgytgkwceyltsisfvhnnsfveleplrtrpeanvtivfssaeqngilmydgq 1194
Qy 420 FTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPL-A-SDGTITMTLRTTSKIGILLYYGD 477

Db 1195 dahlavelfngrirvsydvgnhpvstmysfemvadgkyqavel-laikknftlrvdrgla 1253
Qy 478 DHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCF-LQIDKNPV 536

Db 1254 rsiinegsndylklt-tpm-flgglpvdpaqqayknwqirnltsfkgcmkevwinhklvd 1311

Qy 537 QIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPIN 596

Db 1312 fgnaqrqqkitpgca 1326
Qy 597 LQQALENVNTEQSCS 611

RESULT 2
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 30.8%; Score 1673; DB 30; Length 1534;
Best Local Similarity 38.8%; Pred. No. 1.47e-115;
Matches 255; Conservative 164; Mismatches 203; Indels 36; Gaps 24;

Db 748 snkhlralpkgipknvtelyldgnqftlvpgq-lstfky-lqlvdlsnnkisslsnssft 805
Qy 1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRL-YSLTKLDLSHNRLISLENNTFS 59

Db 806 nmsqlttlilsynalqcipplafqglrslrlllslhgndistlqegifadvtslshlaiga 865
Qy 60 NLTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGS 119

Db 866 nplycdchlrwlsswvktgykepgiarcagpqdmegkllltpakkfecqgpptlavqak 925
Qy 120 NSLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATK 179

Db 926 cdlclsspcqnqgtchndplevyrcacpsgykgrdcevslnscsspgpcenggtchaqege 985
Qy 180 CDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKV--AQ 237

Db 986 dapftcscptgfegptcgvntddcvdhacangg-----v--c---v-d--g--vgnytcq 1030
Qy 238 AGRFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db 1031 cplqyegkaceqlvdlcspdlnpcqheaqcvgtpdgprcecmpgyagdncsenqddcrdh 1090
Qy 298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db 1091 rcqngaqcmdevnsysclcaegysgqlceipphlp-a-pks-pcegtecqngancvd-qg 1146
Qy 358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQG-ECVASQN 416

Db 1147 nrp-vcqclpgfggpecekllsvnfvdrdtylqftdlqnwpranitlqvstaednqilly 1205
Qy 417 SSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPLAS-D-GTITMTLRTTSKIGILLY 474

```
Db   1206 ngdndhiavelyqghvrvsydpgsypssaiysaetindgqfhtvelvafdqmvnlsidgg 1265
          ||:  :: |||:|:|:: |  |::|:| :||:  :|||  | : : : :::  | ||
Qy    475 YGDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKN 534

Db   1266 spmtmdnfgkhytlnsea-p-lyvggmpvdvnsaafrlwqilngtgfhgcirnlyinn-- 1321
          :   ::| ||   | : :   ||:||:|:: :  | |  :: |: :: ||| :: ||:
Qy    535 PVQIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVP 594

Db   1322 -elqd-ftktqmkpgvvpgcepcrklyclhgicqpna-tp-gpmchceagwvglhcdq 1375
           :||: : :    : : : |  : | :| |  || :| | ||:|::   | |||:
Qy    595 INLQQALENVNTEQSCSATVNFCAGIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCDE 652


RESULT   3
ID   W11719 standard; Protein; 727 AA.
AC   W11719;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide.
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          184..228
FT                   /label= DSL
FT   domain          229..261
FT                   /label= EGF1
FT   domain          262..292
FT                   /label= EGF2
FT   domain          293..332
FT                   /label= EGF3
FT   domain          333..370
FT                   /label= EGF4
FT   domain          371..409
FT                   /label= EGF5
FT   domain          410..447
FT                   /label= EGF6
FT   domain          448..485
FT                   /label= EGF7
FT   domain          486..523
FT                   /label= EGF8
FT   domain          524..534
FT                   /label= EGF9
FT   domain          555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58897.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W11719) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
CC   C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
CC   amino acid sequence was deduced from a cDNA clone (T58897) obtd.
CC   from chick stage 4-6 embryos.  An alternatively spliced variant
CC   (W00876) was also isolated, and mouse (W11720) and human (W11721-
CC   38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC   can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
CC   cancer, melanoma or seminoma, and nervous system disorders or to
CC   promote tissue regeneration and repair.
SQ   Sequence   727 AA;

  Query Match           9.8%;  Score 533;  DB 21;  Length 727;
  Best Local Similarity  38.7%;  Pred. No. 1.02e-28;
  Matches   79;  Conservative   40;  Mismatches  68;  Indels  17;  Gaps  12;

Db    302 hkpckngatctntgqgsytcscrpgytgssceieinecdanpckngqsctdlens-ysct 360
          : |||| | |  |:   |||:| ||: |  || :|: | ::|| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    361 cppgfygknce--lsa-m-t-cadgp-cfnggr-ctd---n--p-d-ggyscrcplgysg 406
          |  || |  ||  :   : : | :|  | :  | |::   |  : : :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    407 fncekkidycss--spcangaqcvdlgnsyicqcqagftgrhcddnvddcasfpcvnggt 464
           :|| |::||:   :|| | : |: :  || | | :|||| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    465 cqdgvndysctcppgyngkncstp 488
          | ||:  | | | ||| |  |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT   4
ID   W00876 standard; Protein; 740 AA.
AC   W00876;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide (alternatively spliced variant).
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          184..228
FT                   /label= DSL
FT   domain          229..261
FT                   /label= EGF1
FT   domain          262..292
FT                   /label= EGF2
FT   domain          293..332
FT                   /label= EGF3
FT   domain          333..370
FT                   /label= EGF4
FT   domain          371..409
FT                   /label= EGF5
FT   domain          410..447
FT                   /label= EGF6
FT   domain          448..485
FT                   /label= EGF7
FT   domain          486..523
FT                   /label= EGF8
FT   domain          524..534
FT                   /label= EGF9
FT   domain          555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58898.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W00876) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
CC   C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
```

```
CC   amino acid sequence was deduced from a cDNA clone (T58898) obtd.
CC   from chick stage 4-6 embryos.  A shorter version (W58877) of
CC   C-Delta-1, lacking the 12 C-terminal amino acids of the longer
CC   version, was also isolated, and mouse (W11720) and human (W11721-
CC   38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC   can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
CC   cancer, melanoma or seminoma, and nervous system disorders or to
CC   promote tissue regeneration and repair.
SQ   Sequence  740 AA;

  Query Match           9.8%;  Score 533;  DB 21;  Length 740;
  Best Local Similarity  38.7%;  Pred. No. 1.02e-28;
  Matches   79;  Conservative   40;  Mismatches  68;  Indels  17;  Gaps  12;

Db    302 hkpckngatctntgqgsytcscrpgytgssceieinecdanpcknggsctdlens-ysct 360
          : |||| | |  |:   |||:| ||: |  || :|: | ::|| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    361 cppgfygknce--lsa-m-t-cadgp-cfnggr-ctd---n--p-d-ggyscrcplgysg 406
          |  || |  ||  :   : : | :|  | :  | |::   |  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    407 fncekkidycss--spcangaqcvdlgnsyicqcqagftgrhcddnvddcasfpcvnggt 464
           :|| |::||:   :|| | : |: :  || | | :|||| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    465 cqdgvndysctcppgyngkncstp 488
          | ||:  | | | ||| |  |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT   5
ID   W18348 standard; protein; 520 AA.
AC   W18348;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 3; Page 59-61; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence  520 AA;

  Query Match           9.7%;  Score 525;  DB 25;  Length 520;
  Best Local Similarity  38.7%;  Pred. No. 3.98e-28;
  Matches   79;  Conservative   41;  Mismatches  67;  Indels  17;  Gaps  12;

Db    274 hkpckngatctntgqgsytcscrpgytgatcelgidecdpspcknggsctdlens-ysct 332
          : |||| | |  |:   |||:| ||: |  || :|: | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    333 cppgfygkice--lsa-m-t-cadgp-cfnggr-csd---s--p-d-ggyscrcpvgysg 378
          |  || |  ||  :   : : | :|  | :  | ||:   :  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    379 fncekkidycss--spcsngakcvdlgdaylcrcqagfsgrhcddnvddcasspcanggt 436
           :|| |::||:   :|| | :|: :  :| | | :||:| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    437 crdgvndfsctcppgytgrncsap 460
          | ||:  : | | |||:|: |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT   6
ID   W80813 standard; Protein; 685 AA.
AC   W80813;
DT   16-FEB-1999  (first entry)
DE   Nucleotide sequence of the human Delta 3 protein.
KW   Human; Delta 3 protein; agonist; tissue regeneration;
KW   neurodegenerative disease; neurodifferentiative disorder;
KW   neurodevelopmental disorder; peripheral neuropathy;
KW   spinocerebella degeneration; antagonist; neoplastic disease;
KW   hyperplastic disease; cancer; Waldenstroem's macroglobulemia;
KW   fibroproliferative disorder; cerebravascular tissue; gene therapy;
KW   antibody.
OS   Homo sapiens.
PN   WO9845434-A1.
PD   15-OCT-1998.
PF   06-APR-1998; U06775.
PR   11-JUN-1997; US-872855.
PR   04-APR-1997; US-832633.
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI   Gearing DP, McCarthy SA;
DR   WPI; 98-594482/50.
DR   N-PSDB; V68523.
PT   New isolated human Delta3 gene - used to develop products for
PT   treating, e.g. nerve injury, neurodegenerative disorders, peripheral
PT   neuropathies and spinocerebella degenerations
PS   Claim 2; Fig 1; 160pp; English.
CC   This is the amino acid sequence of the human Delta 3 protein
CC   used in the method of the invention. The Delta3 gene is involved in
CC   the growth and differentiation of cells. Delta3 agonists can be used
CC   for promoting the tissue regeneration or repair needed to treat a
CC   nerve injury, neurodegenerative disease, neurodifferentiative or
CC   neurodevelopmental disorders including peripheral neuropathies and
CC   spinocerebella degenerations. Delta3 antagonists can be used to treat
CC   neoplastic or hyperplastic diseases, e.g. cancers, Waldenstroem's
CC   macroglobulemia and fibroproliferative disorders, particularly of
CC   cerebravascular tissue. The nucleic acids can also be used for gene
CC   therapy. The products can also be used for antibody production,
CC   detection, diagnosis and drug screening.
SQ   Sequence  685 AA;

  Query Match           9.7%;  Score 527;  DB 37;  Length 685;
  Best Local Similarity  37.8%;  Pred. No. 2.83e-28;
  Matches   95;  Conservative   45;  Mismatches  87;  Indels  24;  Gaps  11;

Db    291 hspckngatcsnsgqrsytctcrpgytgvdcelelsecdsnpcrnggsckdqedg-yhcl 349
          :||||| | |  :: | ||| | ||: || || ::  | ::|| | ::||  : | ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    350 cppgyyglhcehstlscadspcfnggsc----r----e-rn-qg--an-yacecppnftg 396
          |  |: |  || :   | :| | ||| |    |    | :| |:   | | |:|| :: |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    397 sncekkvdrcts--npcanggqclnrgpsrmcrcrpgftgtycelhvsdcarnpcahggt 454
           :|| |:: ||   ||| | | |:    |  | | ||||  || :: ||    | :||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    455 chdlenglmctcpagfsgrrcevrt--si-----dacasspcfnratcytdlstdtfvcn 507
          | |   :  | | :|::|: ||:     :     ||| |:|  :: | :  ::  | |:
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSAC-GQGECVASQNSSDFTCK 423

Db    508 cpygfvgsrce 518
          |  || |: |:
Qy    424 CHEGFSGPSCD 434
```

```
RESULT    7
ID   W18349 standard; protein; 702 AA.
AC   W18349;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 4; Page 61-64; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   702 AA;

  Query Match           9.7%;  Score 525;  DB 25;  Length 702;
  Best Local Similarity  38.7%;  Pred. No. 3.98e-28;
  Matches   79;  Conservative   41;  Mismatches  67;  Indels  17;  Gaps  12;

Db    274 hkpckngatctntgqgsytcscrpgytgatcelgidecdpspcknggsctdlens-ysct 332
          : |||| | |  |:   |||:| ||: |  ||  || | |||  | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    333 cppgfygkice--lsa-m-t-cadgp-cfnggr-csd---s--p-d-ggyscrcpvgysg 378
          |  || |  ||  :   : : | :|  | :  | ||:   :  : : :|  | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    379 fncekkidycss--spcsngakcvdlgdaylcrcqagfsgrhcddnvddcasspcanggt 436
           :|| |::||:   :|| | :||: :  :| | | :||:| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

      437 crdgvndfsctcppgytgrncsap 460
          | ||:  : | | |||:|: |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT    8
ID   W11720 standard; Protein; 722 AA.
AC   W11720;
DT   28-APR-1997  (first entry)
DE   M-Delta-1 polypeptide.
KW   M-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Mus sp.
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58899.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Claim 4; Fig 8; 135pp; English.
CC   M-delta-1 polypeptide (W11720) is the mouse homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  It is expressed
CC   primarily in presomitic mesoderm, the central and peripheral
CC   nervous systems, and kidney.  Chick (W11719) and human (W11721-
CC   38) Delta-1 polypeptides have also been identified.  Delta-1
CC   proteins can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
CC   cancer, melanoma or seminoma, as well as nervous system disorders,
CC   and to promote tissue regeneration and repair.
SQ   Sequence   722 AA;

  Query Match           9.7%;  Score 527;  DB 21;  Length 722;
  Best Local Similarity  35.6%;  Pred. No. 2.83e-28;
  Matches   89;  Conservative   54;  Mismatches  80;  Indels  27;  Gaps  20;

Db    294 hkpcrngatctntgqgsytcscrpgytgancelevdecapspckngasctdleds-fsct 352
          : ||:| | |  |:   |||:| ||: | :|| ::| |  ||| | |:|   : : |:|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    353 cppgfygkvce--lsa-m-t-cadgp-cfnggr-csd---n--p-d-ggytchcplgfsg 398
          |  || |  ||  :   : : | :|  | :  | ||:   |  : : :| | ||: : |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    399 fncekkmdlcgs--spcsngakcvdlgnsylcrcqagfsgrycednvddcasspcanggt 456
           :|| |:: |    :|| | :||: :  || | | :||:|  || |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    457 crdsvndfsctcppgytgkncs-apv-sr-ceha-pch-n--g-atc--hqrgqrymcec 506
          | |::  : | | |||:|  |  :|:    : : :|: :  | : |   | :  : | |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKC 424

Db    507 aqgyggpncq 516
           :|::||:|:
Qy    425 HEGFSGPSCD 434


RESULT    9
ID   W18353 standard; protein; 723 AA.
AC   W18353;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Signal
FT   Protein         22..723
FT                   /label= Differentiation_suppression_protein
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
DR   N-PSDB; T70174.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 15; Page 77-82; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   723 AA;
```

```
  Query Match           9.7%;  Score 525;  DB 25;  Length 723;
  Best Local Similarity  38.7%;  Pred. No. 3.98e-28;
  Matches   79;  Conservative   41;  Mismatches  67;  Indels  17;  Gaps  12;

Db    295 hkpckngatctntgqgsytcscrpgytgatcelgidecdpspcknggsctdlens-ysct 353
          : |||| | |  |:   |||:| ||: |  ||  || |  ||| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    354 cppgfygkice--lsa-m-t-cadgp-cfnggr-csd---s--p-d-ggyscrcpvgysg 399
          |  || |  ||  :   : : | :|  | :  | ||:   :  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    400 fncekkidycss--spcsngakcvdlgdaylcrcqagfsgrhcddnvddcasspcanggt 457
           :|| |::||:   :|| | :||: :  :| | | :||:| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    458 crdgvndfsctcppgytgrncsap 481
          | ||:  : | | |||:|: |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


ULT   10
  R28960 standard; Protein; 833 AA.
AC   R28960;
DT   01-APR-1993  (first entry)
DE   Delta D11.
KW   Human; Notch; plasmid; cDNA; clone; D11; expression library; PCR;
KW   polymerase chain reaction; primer; cloning vector; Delta; Serrate;
KW   neurogenic; toporythmic; homotypic; heterotypic; differentiation;
KW   quantitation; antibody.
OS   Homo sapiens.
PN   WO9219734-A.
PD   12-NOV-1992.
PF   01-MAY-1992; U03651.
PR   03-MAY-1991; US-695189.
PR   14-NOV-1991; US-791923.
PA   (INDV ) UNIV INDIANA FOUND.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Blaumueller CM, Fehon RG, Muskavitch MAT;
PI   Rebay I, Shepard SB;
DR   WPI; 92-398861/48.
DR   N-PSDB; Q30997.
PT   Human Notch and Delta DNA and protein sequences - used for study
PT   and manipulation of differentiation processes
PS   Claim 50; Fig 13; 239pp; English.
CC   The sequence given is encoded by the nucleotide sequence of human
CC   Delta gene contained in plasmid cDNA clone D11.  A human expression
CC   library was constructed and screening assays were carried out on to
CC   select for the expressed Delta product.  Alternatively the sequences
CC   could be isolated by amplification using polymerase chain reaction
CC   (PCR) primers.  The isolated gene may be inserted into a cloning
CC   vector and expressed.  The Delta gene and also the Notch and Serrate
CC   neurogenic genes are designated "toporythmic" genes.  The proteins
CC   they encode are involved in specific homo- or heterotypic interactions
CC   crucial to differentiation.  The quantitation of mRNA for human Notch
CC   and Delta and adhesive molecules, and study of its expression are
CC   possible using the DNA and antibodies raised against the Notch and
CC   Delta proteins.
SQ   Sequence   833 AA;

  Query Match           9.6%;  Score 521;  DB 6;  Length 833;
  Best Local Similarity  37.1%;  Pred. No. 7.87e-28;
  Matches   95;  Conservative   55;  Mismatches  77;  Indels  29;  Gaps  22;

Db    295 ctnhrpckngtcfntgeglytckcapgysgddceneiyscdadvnpcqnggtcidepht 354
          | |  |||| : |  |:   |||:|:||: |  |||:| :| :  :|| | :||   :::
Qy    183 CLN-SPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYG--SPCLNNATC-KVAQA 238

Db    355 ktgykchcangwsgkmceekvltcsdkpchqg-icrn-vr--pg-lgs-kg-q--gyqce 405
            ::| | :|  |  || ::  | :  |  |  | : ||  :  | :  : |  :|:|:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db    406 cpigysgpncdlqldncsp--npcinggscqp-sgk--cicpagfsgtrcetniddclgh 460
          ||: | | :|:  |: |:   ||| | | | | :|   |:|::||:|  ||||||||
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db    461 qcenggtcidmvnqyrcqcvpgfhgthc--sskvdl-cl-irpc--an-g-gtcl-nlnn 511
          :|:|||:|:| :  | | | ||: |  |  ::::|:     :|  :  | | |: : |:
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db    512 -dyqctcragftgkdc 526
           |: | |: ||:|  |
Qy    418 SDFTCKCHEGFSGPSC 433


RESULT   11
ID   W18351 standard; protein; 1036 AA.
AC   W18351;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 5; Page 66-71; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1036 AA;

  Query Match           9.0%;  Score 489;  DB 25;  Length 1036;
  Best Local Similarity  36.3%;  Pred. No. 1.81e-25;
  Matches   78;  Conservative   40;  Mismatches  83;  Indels  14;  Gaps  12;

Db    423 cl-gqcqndascrdlvn-gyrcicppgyagdhcerdidecasnpclngghcq-neinrfq 479
          || : | |:| |    :  | | | ||: | ||| :|| | ::|||| : |   :  ||
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    480 clcptgfsgnlcqldidycepnpcqngaqcynrasdyfckcpedyegkncshlkdhcrtt 539
          | |  || |: |: :|| |  : |:||: | :     ||  :|:       :::   |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVR--FC--SEELKNFQSFQINSY-RCD 297

Db    540 pcevidsctvamasndtpegvryissnvcgphgkcksqsggkftcdcnkgftgtycheni 599
           | : :    ::  :|  |       | |  :||| : :|: ::| |: |||| |  ||
Qy    298 -CPM-E-YE-GKHCEDKLEYCT-KKLNPCENNGKCIPINGS-YSCMCSPGFTGNNCETNI 351

Db    600 ndcesnpcrnggtcidgvnsykcicsdgwegayce 634
          :|| :  |:|||:|:||: || |:|  |  | |||
Qy    352 DDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT   12
ID   W18352 standard; protein; 1187 AA.
AC   W18352;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
```

PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 6; Page 71-76; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1187 AA;

Query Match 9.0%; Score 489; DB 25; Length 1187;
Best Local Similarity 36.3%; Pred. No. 1.81e-25;
Matches 78; Conservative 40; Mismatches 83; Indels 14; Gaps 12;

```
Db    423 cl-gqcqndascrdlvn-gyrcicppgyaqdhcerdidecasnpclngghcq-neinrfq 479
          || : | |:| |    :  | | | ||: | ||| :|| | ::|||| : |   :  ||
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    480 clcptgfsgnlcqldidycepnpcqngaqcynrasdyfckcpedyegkncshlkdhcrtt 539
          | |  || |: |: :|| |  : |:||: | :    ||  :|:      :::   |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVR--FC--SEELKNFQSFQINSY-RCD 297

Db    540 pcevidsctvamasndtpegvryissnvcgphgkcksqsggkftcdcnkgftgtycheni 599
            | : :   ::  :|  |      | |  :||| : :|: ::| |: |||| |  ||
Qy    298 -CPM-E-YE-GKHCEDKLEYCT-KKLNPCENNGKCIPINGS-YSCMCSPGFTGNNCETNI 351

Db    600 ndcesnpcrnggtcidgvnsykcicsdgwegayce 634
          :|| :  |:|||:|:||: || |:|  |  | |||
Qy    352 DDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCE 386
```

RESULT 13
ID W44301 standard; Protein; 1218 AA.
AC W44301;
DT 19-JUN-1998 (first entry)
DE Human serrate 1.
KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	Peptide	1..31
FT		/label= Signal
FT	Protein	32..1218
FT		/label= Serrate-1

PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15201.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Disclosure; Page 77-86; 103pp; Japanese.
CC The present sequence represents human serrate 1, from the present
CC invention which describes human serrate 2. The present invention also
CC describes a method for the preparation of the polypeptides, and
CC antibodies binding to the polypeptide and its fragments. The polypeptide
CC and its fragments expressed by the serrate-2-gene can be used to inhibit
CC stem (especially blood stem) cell differentiation and to inhibit
CC endothelial cell growth. They may be incorporated in a cell culture
CC media for culturing undifferentiated stem cells. They can also be used
CC for treatment of neoplasms such as leukaemia. The antibodies can be used
CC for the diagnosis of malignant tumours.
SQ Sequence 1218 AA;

Query Match 9.0%; Score 489; DB 29; Length 1218;
Best Local Similarity 36.3%; Pred. No. 1.81e-25;
Matches 78; Conservative 40; Mismatches 83; Indels 14; Gaps 12;

```
Db    454 cl-gqcqndascrdlvn-gyrcicppgyaqdhcerdidecasnpclngghcq-neinrfq 510
          || : | |:| |    :  | | | ||: | ||| :|| | ::|||| : |   :  ||
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    511 clcptgfsgnlcqldidycepnpcqngaqcynrasdyfckcpedyegkncshlkdhcrtt 570
          | |  || |: |: :|| |  : |:||: | :    ||  :|:      :::   |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVR--FC--SEELKNFQSFQINSY-RCD 297

Db    571 pcevidsctvamasndtpegvryissnvcgphgkcksqsggkftcdcnkgftgtycheni 630
            | : :   ::  :|  |      | |  :||| : :|: ::| |: |||| |  ||
Qy    298 -CPM-E-YE-GKHCEDKLEYCT-KKLNPCENNGKCIPINGS-YSCMCSPGFTGNNCETNI 351

Db    631 ndcesnpcrnggtcidgvnsykcicsdgwegayce 665
          :|| :  |:|||:|:||: || |:|  |  | |||
Qy    352 DDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCE 386
```

RESULT 14
ID W05833 standard; Protein; 1218 AA.
AC W05833;
DT 28-JAN-1997 (first entry)
DE Human Serrate-1 (HJ1).
KW Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
KW cell fate; central nervous system; cancer; tissue repair; therapy;
KW diagnosis; antibody.
OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	domain	1..1067
FT		/label= Extracellular_domain
FT	peptide	14..29
FT		/label= Sig_peptide
FT	domain	185..229
FT		/label= DSL
FT		/note= "region of homology with Drosophila Delta
FT		and Serrate, predicted to mediate binding
FT		with Notch"
FT	domain	234..896
FT		/label= ELR
FT		/note= "epidermal growth factor-like repeat domain"
FT	region	234..264
FT		/label= ELR1
FT	region	265..299
FT		/label= ELR2
FT	region	300..339
FT		/label= ELR3
FT	region	340..377
FT		/label= ELR4
FT	region	378..415
FT		/label= ELR5
FT	region	416..453
FT		/label= ELR6
FT	region	454..490
FT		/label= ELR7
FT	region	491..528
FT		/label= ELR8
FT	region	529..566
FT		/label= ELR9
FT	region	567..598
FT		/label= Partial_ELR
FT	region	599..632
FT		/label= Partial_ELR
FT	region	633..670
FT		/label= ELR10
FT	region	671..708
FT		/label= ELR11
FT	region	709..747

```
FT                   /label= ELR12
FT   region          748..785
FT                   /label= ELR13
FT   region          786..823
FT                   /label= ELR14
FT   region          824..862
FT                   /label= ELR15
FT   region          863..879
FT                   /label= Partial_ELR
FT   region          880..896
FT                   /label= Partial_ELR
FT   domain          1068..1089
FT                   /label= Transmembrane_domain
FT   domain          1090..1218
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI   Lewis JH, Mann RS, Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40090.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Claim 4; Page 95-98; 161pp; English.
CC   Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC   for the zygotic neurogenic locus Notch, and are believed to play a
CC   major role in determining cell fates (differentiation) in the
CC   central nervous system. Their amino acid sequences were deduced
CC   from cDNA clones (see also T40090-91) isolated from human foetal
CC   brain cDNA libraries. The proteins, antibodies raised to them,
CC   and encoding nucleic acids can be used in the detection of
CC   Serrate sequences and in the treatment of disorders of cell fate
CC   or differentiation, partic. cancer, nervous system disorders
CC   and in tissue repair or regeneration.
SQ   Sequence  1218 AA;

  Query Match           9.0%;  Score 489;  DB 19;  Length 1218;
  Best Local Similarity 36.3%;  Pred. No. 1.81e-25;
  Matches   78;  Conservative   40;  Mismatches  83;  Indels  14;  Gaps  12;

Db    454 cl-gqcqndascrdlvn-gyrcicppgyagdhcerdidecasnpclngghcq-neinrfq 510
          || : | |:| |    :  | | | ||: | ||| :|| | ::|||| : |   :  ||
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    511 clcptgfsgnlcqldidycepnpcqngaqcynrasdyfckcpedyegkncshlkdhcrtt 570
          | |  || |: |: :|| |  : |:||: | :     ||  :|:       :::   |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVR--FC--SEELKNFQSFQINSY-RCD 297

Db    571 pcevidsctvamasndtpegvryissnvcgphgkcksqsggkftcdcnkgftgtycheni 630
          | : :   ::  :|  |      | |  :||| : :|: ::| |: |||| |  ||
Qy    298 -CPM-E-YE-GKHCEDKLEYCT-KKLNPCENNGKCIPINGS-YSCMCSPGFTGNNCETNI 351

Db    631 ndcesnpcrnggtcidgvnsykcicsdgwegayce 665
          :|| :  |:|||:|:||: || |:|  |  | |||
Qy    352 DDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT  15
ID   W40827 standard; Protein; 1208 AA.
AC   W40827;
DT   21-MAY-1998  (first entry)
DE   Human Jagged protein.
KW   Jagged; Notch; angiogenesis; endothelial cell; migration; human;
KW   wound repair; vulnerary; injury repair; signal transduction;
KW   motor neurone disease; amyotrophic lateral sclerosis; polymyelitis;
KW   diagnosis; therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..11
FT                   /label= Sig_peptide
FT   Domain          175..220
FT                   /note= "DSL (Delta, Serrate, Lag-2 and Apx-1)
FT                        domain"
FT   Region          224..852
FT                   /note= "EGF-like repeat region containing 16
FT                        EGF repeats"
FT   Misc_difference 526
FT                   /note= "encoded by ANC"
FT   Region          853..992
FT                   /note= "cysteine-rich region"
FT   Domain          1058..1083
FT                   /note= "transmembrane domain"
FT   Region          1084..1208
FT                   /note= "cytoplasmic region"
PN   WO9745143-A1.
PD   04-DEC-1997.
PF   30-MAY-1997; U09407.
PR   31-MAY-1996; US-018841.
PA   (NAAM-) NAT AMERICAN RED CROSS.
PA   (UYGE-) UNIV GENEVE.
PI   Maciag T, Montesano R, Pepper M, Wong MK, Zimrin AB;
DR   WPI; 98-032340/03.
DR   N-PSDB; V03674.
PT   New human Jagged protein - used to inhibit or promote angiogenesis
PT   and to control migration of endothelial cells in injured blood
PT   vessels
PS   Claim 2; Page 54-61; 81pp; English.
CC   This sequence comprises the human homologue of the rat Jagged
CC   protein. Jagged is able to bind Notch protein and is involved in
CC   endothelial cell (EC) migration and differentiation. The human
CC   Jagged amino acid sequence was deduced from a human endothelial
CC   cell cDNA (see V03674) induced by exposure to fibrin. Jagged
CC   polypeptides can be expressed in host cell systems. A method for
CC   treating or preventing disease by administering an agent that
CC   (ant)agonises, inhibits, prevents, enhances or stimulates function
CC   of the Notch or Jagged proteins is claimed, as well as a method for
CC   affecting differentiation of mesoderm, endoderm, ectoderm and/or
CC   neuroderm cells. When Jagged is applied to a micro-diameter blood
CC   vessel from which ECs have been removed, damaged or reduced, it
CC   decrease migrations of EC to the site, but when delivered to a
CC   similar site on a large vessel it increases EC migration. Jagged
CC   and its agonists are used to inhibit or prevent angiogenesis (where
CC   associated with solid tumours, rheumatoid arthritis, inflammation,
CC   or restenosis, particularly preventing angiogenesis from the vaso
CC   vasorum and promoting large vessel EC migration to repair the lumen
CC   of large vessels). Anti-Jagged and Jagged antagonists (e.g.
CC   antisense Jagged and Jagged mutants) are used to promote or enhance
CC   angiogenesis, particularly for wound and injury repair, e.g. where
CC   surgical, traumatic and/or caused by disease, e.g. diabetes-related
CC   (all claimed). Angiogenesis can be modulated in vitro or in vivo
CC   and expression of proteins by gene therapy is included. Modulation
CC   of the Notch-Jagged signalling pathway may also be involved in
CC   placental development and motor neurone diseases such as
CC   amyotrophic lateral sclerosis, poliomyelitis etc.
SQ   Sequence  1208 AA;

  Query Match           8.9%;  Score 484;  DB 28;  Length 1208;
  Best Local Similarity 35.8%;  Pred. No. 4.22e-25;
  Matches   77;  Conservative   40;  Mismatches  84;  Indels  14;  Gaps  12;

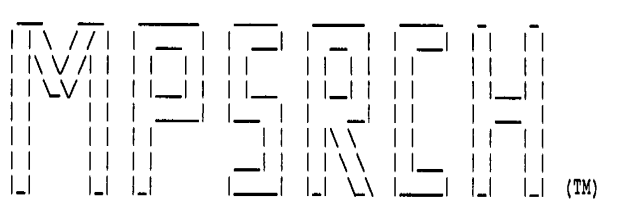
Db    444 cl-gqcqndascrdlvn-gyrcicppgyagdhcerdidecasnpclngghcq-neinrfq 500
          || : | |:| |    :  | | | ||: | ||| :|| | ::|||| : |   :  ||
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    501 clcptgfsgnlcqldidycepnpcqxgaqcynrasdyfckcpedyegkncshlkdhcrtt 560
          | |  || |: |: :|| |  : |: |: | :     ||  :|:       :::   |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVR--FC--SEELKNFQSFQINSY-RCD 297

Db    561 pcevidsctvamasndtpegvryissnvcgphgkcksqsggkftcdcnkgftgtycheni 620
          | : :   ::  :|  |      | |  :||| : :|: ::| |: |||| |  ||
Qy    298 -CPM-E-YE-GKHCEDKLEYCT-KKLNPCENNGKCIPINGS-YSCMCSPGFTGNNCETNI 351
```

```
Db    621 ndcesnpcrnggtcidgvnsykcicsdgwegayce 655
          :|| :  |:|||:|:||: || |:|  |  | |||
Qy    352 DDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCE 386
```

Search completed: Fri May 28 09:09:55 1999
Job time : 115 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:10:14 1999; MasPar time 31.17 Seconds
944.872 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-9
Description: (1-735) from US09191647.pep
Perfect Score: 5438
Sequence: 1 SNKNLTSFPSRIPFDTTELY.....TVHIIRQCQCEPTKSVLSEK 735

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.886; Variance 100.304; scale 0.497

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1894	34.8	1469	2	B36665	slit protein 2 precur	0.00e+00
2	1838	33.8	1480	2	A36665	slit protein 1 precur	0.00e+00
3	1188	21.8	530	2	A31640	epidermal growth fact	9.23e-210
4	588	10.8	1064	2	A40136	fibropellin Ia - sea	2.04e-87
5	559	10.3	570	2	A48836	fibropellin C precurs	1.13e-81
6	557	10.2	2703	2	A24420	notch protein - fruit	2.80e-81
7	548	10.1	2437	2	S42612	transmembrane protein	1.67e-79
8	538	9.9	861	2	A48825	Notch homolog Motch p	1.56e-77
9	540	9.9	1203	2	A49175	Motch B protein - mou	6.30e-78
10	540	9.9	2524	2	A35844	Xotch protein - Afric	6.30e-78
11	533	9.8	728	2	I50719	C-Delta-1 - chicken	1.50e-76
12	527	9.7	722	2	I48324	DELTA-like 1 - mouse	2.27e-75
13	529	9.7	2139	2	A35672	crumbs protein - frui	9.19e-76
14	527	9.7	2471	2	A49128	cell-fate determining	2.27e-75
15	520	9.6	832	2	A31246	neurogenic protein De	5.38e-74
16	521	9.6	833	2	S19087	gene Delta protein pr	3.42e-74
17	520	9.6	880	2	S00670	gene Delta protein pr	5.38e-74
18	520	9.6	2531	2	S18188	notch protein homolog	5.38e-74
19	518	9.5	2531	2	A46019	gene Notch-1 protein	1.33e-73
20	505	9.3	2318	2	S45306	notch 3 protein - mou	4.67e-71
21	508	9.3	2555	2	A40043	notch protein homolog	1.21e-71
22	486	8.9	2321	2	S78549	notch3 protein - huma	2.39e-67
23	472	8.7	293	2	B26637	neurogenic repetitive	1.26e-64
24	474	8.7	1220	2	A56136	jagged protein precur	5.16e-65
25	434	8.0	1429	2	S06434	homeotic protein lin-	2.79e-57
26	417	7.7	383	2	B45484	delta-like dlk homeot	5.08e-54
27	417	7.7	383	2	S53716	homeotic protein dlk	5.08e-54
28	416	7.6	259	2	S48713	fetal antigen 1 - hum	7.90e-54
29	412	7.6	1404	2	A36666	serrate protein precu	4.59e-53
30	412	7.6	1408	2	S16148	gene serrate protein	4.59e-53
31	406	7.5	260	2	A44549	fetal antigen 1 homeo	6.42e-52
32	401	7.4	385	2	S53718	homeotic protein dlk	5.75e-51
33	391	7.2	385	2	A54785	preadipocyte factor 1	4.57e-49
34	391	7.2	1295	2	A32901	glp1 protein precurso	4.57e-49
35	384	7.1	387	2	B49175	Motch A protein - mou	9.69e-48
36	380	7.0	200	2	A26637	neurogenic repetitive	5.53e-47
37	380	7.0	473	2	A56175	adhesive plaque prote	5.53e-47
38	318	5.8	5147	1	IJFFTM	cadherin-related tumo	2.02e-35
39	307	5.6	102	2	B55885	chondroitin sulfate p	2.09e-33
40	304	5.6	2397	2	A55535	versican precursor -	7.35e-33
41	307	5.6	2409	2	A60979	versican precursor -	2.09e-33
42	300	5.5	862	2	S43922	versican - pig-tailed	3.93e-32
43	292	5.4	3562	2	A47171	chondroitin sulfate p	1.11e-30
44	286	5.3	1091	2	A58532	glial cell membrane g	1.34e-29
45	285	5.2	886	2	A57172	probable hormone rece	2.03e-29

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                homology #label LRR7\
  371-394       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LRR8\
  395-418       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LRR9\
  419-442       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR10\
  450-494       #domain proteoglycan carboxyl-terminal homology #label
                PCS2\
  512-537       #domain proteoglycan amino-terminal homology #label
                PAH3\
  547-571       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR11\
  572-595       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR12\
  596-619       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR13\
  620-643       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR14\
  651-695       #domain proteoglycan carboxyl-terminal homology #label
                PCS3\
  708-733       #domain proteoglycan amino-terminal homology #label
                PAH4\
  743-766       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR15\
  767-790       #domain leucine-rich alpha-2-glycoprotein repeat
                homology #label LR16\
  846-890       #domain proteoglycan carboxyl-terminal homology #label
                PCS4\
  1028-1061     #domain EGF homology #label EGF
SUMMARY       #length 1469  #molecular-weight 164695  #checksum 8361

 Query Match          34.8%;  Score 1894;  DB 2;  Length 1469;
 Best Local Similarity  40.7%;  Pred. No. 0.00e+00;
 Matches  307;  Conservative  159;  Mismatches 240;  Indels  49;  Gaps  28;

Db   729 SRNQLKEIPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFAN 788
         | : |  :|  || :|:||||::| |::|    : :| |||:||||:|::  | | ||:|
Qy     1 SNKNLTSFPSRIPPDTTELYLDANYINEIPAHDLNRLYSLTKLDLSHNRLISLENNTFSN 60

Db   789 LTKLSTLIISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSN 848
         ||:|||||||||||:|||  |::||| ||::|||| || ||:::| :| |:||||:|||
Qy    61 LTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGSN 120

Db   849 PLYCDCGLKWPSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKC 908
         :||||| : ||| |||  ::|:|||||  |: : : |:|:: :  | | ::|   : :||
Qy   121 SLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKC 180

Db   909 NACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLEEGR 968
         : |:: || | | | : : | | | | ||: | |||  ||||||:|| ||||| | : ||
Qy   181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db   969 FSCQCAPGYTGARCETNIDDCL-G--EI--KCQNNAT-C---I-D--G--VESYKCECQP 1014
         |:| |  |: |  || |||||: :  |   || :    |   : : :  ::||:|:|
Qy   241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db  1015 GFSGEFCDTKIQFCSPEFNPCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQ 1074
           : |  |: |:::|:  :||| | :||:     ||| | :|| | ||  ||||| |  ||
Qy   301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db  1075 NGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSD 1134
         |||:|||||  |:| |   |:| |||   |: | |  | :||   | :| |  ::  :||
Qy   361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECV-ASQNSSD 419

Db  1135 YLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQ 1194
         : |:|| |::|  |:   |::| :  ::: |:|| : ::: :|: : ::   |||:| |:
Qy   420 FTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPL-A-SDGTITMTLRTTSKIGILLYYGD 477

Db  1195 DAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVEL-LAIKKNFTLRVDRGLA 1253
         |  :: ||::||::: | :|| | | |||  :| ||  | : :  : :| | |::|:
Qy   478 DHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCF-LQIDKNPV 536

Db  1254 RSIINEGSNDYLKLT-TPM-FLGGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVD 1311

          : : | | :| |     | ::||||:: :|:| :  :::|  |:|||:  : ||   |
Qy   537 QIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINE--VP 594

Db  1312 FGNAQRQQKITPGCALLEGEQQEEEDDEQDFMDETPHIKEEPVDPCLENKCRRGSRCVPN 1371
         : | |  |      ||      |: : ||     | ::    :| |  |    | :|  |
Qy   595 I-NLQ--Q------AL------ENVNTEQSC-SATVNFCAG-ID-C-GN----G-KCTNN 630

Db  1372 SNARDGYQCKCKHGQRGRYCDQAASTCRKEQVREYYTENDCRSRQPLKYAKCVGGCG-NQ 1430
         : :  || | |     |  ||:    | |:  |    ||:||| : :| | | | || :|
Qy   631 ALSPKGYMCQCDSHFSGEHCDEKRIKCDKQKFRRHHIENECRSVDRIKIAECNGYCGGEQ 690

Db  1431 -CCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGC 1464
           ||:|   ::|||:|:| |    |  : |:| | |
Qy   691 NCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQC 725


RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
  66-91             #domain proteoglycan amino-terminal homology #label
                    PAH1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR1\
  125-148           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR2\
  149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR3\
  173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR4\
  197-220           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR5\
  228-272           #domain proteoglycan carboxyl-terminal homology #label
                    PCS1\
  288-313           #domain proteoglycan amino-terminal homology #label
                    PAH2\
  323-346           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR6\
  347-370           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR7\
  371-394           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR8\
  395-418           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR9\
  419-442           #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR10\
  450-494           #domain proteoglycan carboxyl-terminal homology #label
                    PCS2\
```

```
  512-537          #domain proteoglycan amino-terminal homology #label
                     PAH3\
  547-571          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR11\
  572-595          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR12\
  596-619          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR13\
  620-643          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR14\
  651-695          #domain proteoglycan carboxyl-terminal homology #label
                     PCS3\
  708-733          #domain proteoglycan amino-terminal homology #label
                     PAH4\
  743-766          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR15\
  767-790          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR16\
  791-814          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR17\
  815-838          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR18\
  846-890          #domain proteoglycan carboxyl-terminal homology #label
                     PCS4\
  1028-1061        #domain EGF homology #label EGF
SUMMARY          #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           33.8%;  Score 1838;  DB 2;  Length 1480;
  Best Local Similarity  41.8%;  Pred. No. 0.00e+00;
  Matches   257;  Conservative  143;  Mismatches 194;  Indels  21;  Gaps  15;

Db    729 SRNQLKEIPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFAN 788
Qy      1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRLYSLTKLDLSHNRLISLENNTFSN 60

Db    789 LTKLSTLIISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSN 848
Qy     61 LTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGSN 120

Db    849 PLYCDCGLKWFSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKC 908
Qy    121 SLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKC 180

Db    909 NACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLEEGR 968
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    969 FSCQCAPGYTGARCETNIDDCL-G--EI--KCQNNAT-C---I-D--G--VESYKCECQP 1014
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db   1015 GFSGEFCDTKIQFCSPEFNPCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQ 1074
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db   1075 NGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSD 1134
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECV-ASQNSSD 419

Db   1135 YLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQ 1194
Qy    420 FTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPL-A-SDGTITMTLRTTSKIGILLYYGD 477

Db   1195 DAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVEL-LAIKKNFTLRVDRGLA 1253
Qy    478 DHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCF-LQIDKNPV 536

Db   1254 RSIINEGSNDYLKLT-TPM-FLGGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVD 1311
Qy    537 QIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPIN 596

Db   1312 FGNAQRQQKITPGCA 1326
```

```
Qy    597 LQQALENVNTEQSCS 611


RESULT    3
ENTRY           A31640     #type fragment
TITLE           epidermal growth factor-like protein slit - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A31640
REFERENCE       A31640
  #authors      Rothberg, J.M.; Hartley, D.A.; Walther, Z.;
                  Artavanis-Tsakonas, S.
  #journal      Cell (1988) 55:1047-1059
  #title        slit: An EGF-homologous locus of D. melanogaster involved in
                  the development of the embryonic central nervous system.
  #cross-references MUID:89077533
  #accession    A31640
    ##molecule_type DNA
    ##residues      1-530 ##label ROT
    ##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
  #introns      470/3
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        growth-factor
FEATURE
  148-181            #domain EGF homology #label EGF
SUMMARY         #length 530  #checksum 6330

  Query Match           21.8%;  Score 1188;  DB 2;  Length 530;
  Best Local Similarity  36.9%;  Pred. No. 9.23e-210;
  Matches   171;  Conservative  111;  Mismatches 159;  Indels  22;  Gaps  16;

Db      1 MKDKLILSTPSSSFVCRGRVRNDILAKCNACPEQPCQNQAQCVALPQREYQCLCQPGYHG 60
Qy    153 VSNQLLLTAQPYQFTCDSKVPTKLATKCDLCLNSPCKNNAICETTSSRKYTCNCTPGFYG 212

Db     61 KHCEFMIDACYGNPCRNNATCTVLEEGRFSCQCAPGYTGARCETNIDDCL-G--EI--KC 115
Qy    213 VHCENQIDACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKC 272

Db    116 QNNAT-C---I-D--G--VESYKCECQPGFSGEFCDTKIQFCSPEFNPCANGAKCMDHFT 166
Qy    273 VDLVRFCSEELKNFQSFQINSYRCDCPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPING 332

Db    167 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 226
Qy    333 SYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMD 392

Db    227 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 286
Qy    393 MEYQKTDACQQSACGQGECV-ASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALD 451

Db    287 PLRTRPEANVTIVF-SSGQNGILMYDGQDAHLAVELFNGRIRVSYDVGNHPVSTMYSFEM 345
Qy    452 PL-A-SDGTITMTLRTTSKIGILLYYGDDHFVSAELYDGRVKLVYYIGNFPASHMYSSVK 509

Db    346 VADGKYHAVEL-LAIKKNFTLRVDRGLARSIINEGSNDYLKLT-TPM-FLGGLPVDPAQQ 402
Qy    510 VNDGLPHRISIRTSERKCF-LQIDKNPVQIVENSGKSDQLITKGKEMLYIGGLPIEKSQD 568

Db    403 AYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQKITPGCA 445
Qy    569 AKRRFHVKNSESLKGCISSITINEVPINLQQALENVNTEQSCS 611


RESULT    4
ENTRY           A40136     #type complete
TITLE           fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
ALTERNATE_NAMES epidermal growth factor homolog precursor
```

```
CONTAINS        alternatively spliced fibropellin Ib (EGFI)
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            13-May-1992 #sequence_revision 17-Sep-1997 #text_change
                  07-Aug-1998
ACCESSIONS      A40136; B40136; C40136; A29316; A43131
REFERENCE       A40136
  #authors      Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff,
                  R.A.
  #journal      J. Mol. Evol. (1989) 29:314-327
  #title        Structural analysis of the uEGF gene in the sea urchin
                  Strongylocentrotus purpuratus reveals more similarity to
                  vertebrate than to invertebrate genes with EGF-like
                  repeats.
  #cross-references MUID:90112459
  #accession    A40136
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-114 ##label DEL
    ##cross-references GB:X17530; NID:g10225; PID:g667061
  #accession    B40136
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type DNA
    ##residues     181-251,329-370,'R',372-408,'RA',411-441 ##label DE2
  #accession    C40136
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type DNA
    ##residues     'K',747-821,898-978 ##label DE3
REFERENCE       A29316
  #authors      Hursh, D.A.; Andrews, M.E.; Raff, R.A.
  #journal      Science (1987) 237:1487-1490
  #title        A sea urchin gene encodes a polypeptide homologous to
                  epidermal growth factor.
  #cross-references MUID:87319677
  #accession    A29316
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     'S',280-481,786-1064 ##label HUR
    ##cross-references GB:M17421; NID:g161474; PID:g552260
REFERENCE       A43131
  #authors      Hunt, L.T.; Barker, W.C.
  #journal      FASEB J. (1989) 3:1760-1764
  #title        Avidin-like domain in an epidermal growth factor homolog from
                  a sea urchin.
  #cross-references MUID:89196806
  #contents     annotation
COMMENT     EGF homology repeats 10-17 are spliced out in the short form
              (fibropellin Ib).
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
  1-19            #domain signal sequence #status predicted #label SIG\
  20-1064         #product fibropellin I #status predicted #label FIB\
  23-54           #domain EGF homology #label EG01\
  57-175          #domain C1r/C1s repeat homology #label CSR\
  180-211         #domain EGF homology #label EG02\
  218-249         #domain EGF homology #label EG03\
  256-287         #domain EGF homology #label EG04\
  294-325         #domain EGF homology #label EG05\
  332-363         #domain EGF homology #label EG06\
  370-401         #domain EGF homology #label EG07\
  408-439         #domain EGF homology #label EG08\
  446-477         #domain EGF homology #label EG09\
  484-515         #domain EGF homology #label EG10\
  522-553         #domain EGF homology #label EG11\
  560-591         #domain EGF homology #label EG12\
  598-629         #domain EGF homology #label EG13\
  636-667         #domain EGF homology #label EG14\
  674-705         #domain EGF homology #label EG15\
  712-743         #domain EGF homology #label EG16\
  750-781         #domain EGF homology #label EG17\
  788-819         #domain EGF homology #label EG18\
  826-857         #domain EGF homology #label EG19\
  864-895         #domain EGF homology #label EG20\
  902-933         #domain EGF homology #label EG21\
  936-1064        #region avidin-like\
  23-34,28-43,45-54,
  62-88,180-191,
  185-200,202-211,
  218-229,223-238,
  240-249,256-267,
  261-276,278-287,
  294-305,299-314,
  316-325,332-343,
  337-352,354-363,
  370-381,375-390,
  392-401,408-419,
  413-428,430-439,
  446-457,451-466,
  468-477,484-495   #disulfide_bonds #status predicted\
  489-504,506-515,
  522-533,527-542,
  544-553,560-571,
  565-580,582-591,
  598-609,603-618,
  620-629,636-647,
  641-656,658-667,
  674-685,679-694,
  696-705,712-723,
  717-732,734-743,
  750-761,755-770,
  772-781,788-799,
  793-808,810-819,
  826-837,831-846,
  848-857,864-875,
  869-884,886-895,
  902-913,907-922,
  924-933         #disulfide_bonds #status predicted
SUMMARY         #length 1064  #molecular-weight 112072  #checksum 303

  Query Match          10.8%;  Score 588;  DB 2;  Length 1064;
  Best Local Similarity  39.7%;  Pred. No. 2.04e-87;
  Matches  102;  Conservative  47;  Mismatches  79;  Indels  29;  Gaps  17;

Db    216 DECASDPCQNGGAC-VDGVNGYVCNCVPGFDGDECENNINECASSPCLNGGIC-VDGVNM 273
          | | : || | : |   :  | ||| ||| |  ||| |: | :||||| : | |
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    274 FECTCLAGFTGVRCEVNIDECASAPCQNGG-----I--C---I-D--G--INGYTCSCPL 318
          |:| |  || |  || |||:| :: |:|||    :  |   : :  :  ||:| | ||:
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    319 GFSGDNCENNDDECSS-I-PCLNGGTCVDLVNAYMCVCAPGWTGPTCADNIDECASAPCQ 376
           : | :||:: : |:  : || | | |: :  :| |:|:|| ||  |  |||:| :  ||
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    377 NGGVCIDGVNGYMCDCQPGYTGTHCETD--ID-ECARP-PCQ-N--G-GDCVDGVNG--Y 426
          ||| |:||: :| | |:|||:|  ||    :| |  :  :|| :  | |:|| : |:  :
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDF 420

Db    427 VCICAPGFDGLNCENNI 443
           | |  || | :|:  :
Qy    421 TCKCHEGFSGPSCDRQM 437


RESULT    5
ENTRY           A48836     #type complete
TITLE           fibropellin C precursor - sea urchin (Strongylocentrotus
                  purpuratus)
ALTERNATE_NAMES EGF repeat-containing protein; epidermal growth
                  factor-related protein 3; fibropellin III
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  07-Aug-1998
ACCESSIONS      A48836
REFERENCE       A48836
```

```
   #authors      Bisgrove, B.W.; Raff, R.A.
   #journal      Dev. Biol. (1993) 157:526-538
   #title        The SpEGF III gene encodes a member of the fibropellins: EGF
                   repeat-containing proteins that form the apical lamina of
                   the sea urchin embryo.
   #cross-references MUID:93273088
   #accession    A48836
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-570 ##label BIS
      ##cross-references GB:L07045; NID:g310659; PID:g310660
      ##note          sequence extracted from NCBI backbone (NCBIN:132724,
                       NCBIP:132725)
CLASSIFICATION   #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-18             #domain signal sequence #status predicted #label SIG\
   19-570           #product fibropellin C #status predicted #label FIB\
   19-54            #domain EGF homology #label EGF1\
   57-175           #domain C1r/C1s repeat homology #label C1R\
   176-211          #domain EGF homology #label EGF2\
   214-249          #domain EGF homology #label EGF3\
   252-287          #domain EGF homology #label EGF4\
   290-325          #domain EGF homology #label EGF5\
   328-363          #domain EGF homology #label EGF6\
   366-401          #domain EGF homology #label EGF7\
   404-439          #domain EGF homology #label EGF8\
   442-570          #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439    #disulfide_bonds #status predicted
SUMMARY          #length 570  #molecular-weight 61115  #checksum 5567

  Query Match           10.3%;  Score 559;  DB 2;  Length 570;
  Best Local Similarity  42.2%;  Pred. No. 1.13e-81;
  Matches    86;  Conservative   35;  Mismatches  64;  Indels  19;  Gaps   8;

Db    256 CASIPCLNGGIC-VDGINQFACTCLPGYTGILCETDINECASSPCQNGGSCTDA-VNRYT 313
          | : || | :||   :   ::| | ||: |: || :|: | :||| | ::|  |   |:
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSC--L------D-----G--VDGYVCQCLPNY 358
          | |  || |  || ||::| :| | ||| |  |      :     :  :::| |:|  :|
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    359 TGTHCEISLDACAS-L-PCQNGGVCTNVGGDYVCECLPGYTGINCEIDINECASLPCQNG 416
          | |||  |: |:  | ||:| | |  : | | | | ||:|| ||| :|::| :: ||||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    417 GECINGIAMYICQCRQGYAGVNCE 440
          | |::||  | | || ||||  ||
Qy    363 GSCVDGILSYDCLCRPGYAGQYCE 386


RESULT    6
ENTRY            A24420     #type complete
TITLE            notch protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  neurogenic repetitive locus protein
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
                   07-Aug-1998
ACCESSIONS       A24420; A24768; S09358; A05267
REFERENCE        A24420
   #authors      Kidd, S.; Kelley, M.R.; Young, M.W.
   #journal      Mol. Cell. Biol. (1986) 6:3094-3108
   #cross-references MUID:87064624
   #accession    A24420
      ##molecule_type DNA
      ##residues      1-2703 ##label KID
      ##cross-references GB:K03508; NID:g157991; PID:g157993
REFERENCE        A24768
   #authors      Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
   #journal      Cell (1985) 43:567-581
   #cross-references MUID:86079539
   #accession    A24768
      ##molecule_type mRNA
      ##residues      1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',
                       268-872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',
                       2258-2264,'V',2266-2406,'R',2408-2444,'L',2446-2703
                       ##label WHA1
      ##note          the authors translated the codon ATC for residue 49 as
                       Thr, ATT for residue 2044 as Arg, GTA for residue 2265
                       as Ala, CGC for residue 2407 as His, and CTT for
                       residue 2445 as Arg
REFERENCE        S09358
   #authors      Tautz, D.
   #journal      Nucleic Acids Res. (1989) 17:6463-6471
   #title        Hypervariability of simple sequences as a general source for
                   polymorphic DNA markers.
   #cross-references MUID:89385974
   #accession    S09358
      ##molecule_type DNA
      ##residues      2505-2551,'QQQQ',2552-2576,'E',2578-2604 ##label TAU
REFERENCE        A05267
   #authors      Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.;
                   Artavanis-Tsakonas, S.
   #journal      Cell (1985) 40:55-62
   #title        opa: a novel family of transcribed repeats shared by the
                   Notch locus and other developmentally regulated loci in D.
                   melanogaster.
   #cross-references MUID:85099329
   #accession    A05267
      ##molecule_type DNA
      ##residues      2504-2576,'E',2578-2611 ##label WHA2
GENETICS
   #gene         notch; opa
      ##cross-references FlyBase:FBgn0004647
   #map_position 8.96-9.36
   #introns      53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
CLASSIFICATION   #superfamily notch protein; ankyrin repeat homology; EGF
                   homology
KEYWORDS         differentiation; tandem repeat; transmembrane protein
FEATURE
   27-43            #domain transmembrane #status predicted #label TMM1\
   568-599          #domain EGF homology #label EGF\
   1746-1762        #domain transmembrane #status predicted #label TMM2\
   1950-1982        #domain ankyrin repeat homology #label AN1\
   1983-2015        #domain ankyrin repeat homology #label AN2\
   1988-2004        #domain transmembrane #status predicted #label TMM3\
   2017-2049        #domain ankyrin repeat homology #label AN3\
   2050-2082        #domain ankyrin repeat homology #label AN4\
   2083-2115        #domain ankyrin repeat homology #label AN5\
   2538-2568        #region glutamine-rich\
   2538-2568        #domain neurogenic repetitive element #status predicted
                      #label OPA
SUMMARY          #length 2703  #molecular-weight 288876  #checksum 6404

  Query Match           10.2%;  Score 557;  DB 2;  Length 2703;
  Best Local Similarity  38.1%;  Pred. No. 2.80e-81;
  Matches    96;  Conservative   49;  Mismatches  80;  Indels  27;  Gaps  17;

Db    681 CHSNPCNNGATC-IDGINSYKCQCVPGFTGQHCEKNVDECISSPCANNGVC-IDQVNGYK 738
          | ::||:| | |   :   | | | ||| | |||: :| | :||| ||: | : |   ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    739 CECPRGFYDAHCLSDVDECA-S---NP--CVNEGR-CEDGI-N-E-F-I----CHCPPGY 783
          | | :||   | ::|:| |   |  ||:  | | : : | : | |    | ||  |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302
```

```
Db    784 TGKRCELDIDECSS--NPCQHGGTCYDKLNAFSCQCMPGYTGQKCETNIDDCVTNPCGNG 841
          ||:||  :: |:   |||:: | |     ::|| | ||:|| :||||||||    | ||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    842 GTCIDKVNGYKCVCKVPFTGRDCE--SKMD-PCA-SNRCK-N---EAKCTPSSNFLDFSC 893
          |:|:| : :| |:|:  ::|: ||  ::||     :: |  :   :: | :| |  ||:|
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    894 TCKLGYTGRYCD 905
          |  |::|  ||
Qy    423 KCHEGFSGPSCD 434


RESULT    7
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
   #authors     Bierkamp, C.; Campos-Ortega, J.A.
   #journal     Mech. Dev. (1993) 43:87-100
   #title       A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
   #cross-references MUID:94128602
   #accession   S42612
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2437 ##label BIE
      ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1915-1947          #domain ankyrin repeat homology #label AN1\
   1948-1980          #domain ankyrin repeat homology #label AN2\
   1982-2014          #domain ankyrin repeat homology #label AN3\
   2015-2047          #domain ankyrin repeat homology #label AN4\
   2048-2080          #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           10.1%;  Score 548;  DB 2;  Length 2437;
  Best Local Similarity  39.6%;  Pred. No. 1.67e-79;
  Matches  105;  Conservative   56;  Mismatches  70;  Indels  34;  Gaps  24;

Db    528 DECASTPCKNGAKC-TDGPNKYTCECTPGFSGIHCELDINECASSPC-HYGVCRDGVAS- 584
          | | ::|||| | | | :: ||||:||||| |:||| :|: | :||| : : |: : |:
      181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

      585 FTCDCRPGYTGRLCETNINECLSQPCRNGGTCQD--R----E--N--AY-I----CTCPK 629
          | | |  |: |  || ||::|::  | ||| | |  |    |  |  :: |    | ||:
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    630 GTTGVNCEINIDDC-KR-KPCDY-GKCIDKING-YECVCEPGYSGSMCNINIDDCALNPC 685
            | :|| ::: | |: :||:  |||| ||| | |:| ||::|: |: |||||    |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIP-INGSYSCMCSPGFTGNNCETNIDDCKNVEC 359

Db    686 HNGGTCIDGVNSFTCLC-PD--G-FRDATCL-S-Q-HN-E-CSSNPCIHGSCLD-QINS- 734
          :|||:|:||: |: ||| |   | : :   :   : :: : |  ::| :| |:  | :|
Qy    360 QNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSD 419

Db    735 YRCVCEAGWMGRNCDININECLSNP 759
          : | |  |  | :||  ::  : ||
Qy    420 FTCKCHEGFSGPSCDRQMSVGFKNP 444


RESULT    8
ENTRY           A48825     #type fragment
TITLE           Notch homolog Motch protein - mouse (fragment)
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A48825
REFERENCE       A48825
   #authors     Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.;
                  Rossant, J.
   #journal     Dev. Biol. (1992) 154:377-387
   #title       Expression analysis of a Notch homologue in the mouse embryo.
   #cross-references MUID:93050801
   #accession   A48825
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-861 ##label REA
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:119144)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   26-57              #domain EGF homology #label EGF
SUMMARY         #length 861  #checksum 7963

  Query Match           9.9%;  Score 538;  DB 2;  Length 861;
  Best Local Similarity  37.3%;  Pred. No. 1.56e-77;
  Matches   78;  Conservative   39;  Mismatches  83;  Indels   9;  Gaps   7;

Db     26 CESNPCVNGGTCKDMTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTC-IDDVAGYK 83
          | ::|| | : |   :|  | | |  || | :|:  |: | ::|||| :|| : : : ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db     84 CNCPLPYTGATCEVVLAPCATSPCKNSGVCKESEDYESPSCVPCHNGGSCTDGINTAFCD 143
          | |   : |  ||  :  |  | | |:| |    |   | |    :     : ||:  ||
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKC--V-DLVRF-CSEELKNFQSFQ-INSYRCD 297

Db    144 SVPGFQGAFCEEDINECAS--NPCQNGANCTDCVDSYTCTCPVGFNGIHCENNTPDCTES 201
            ::|  ||: :: |:   |||:| ::|     ||:| |: || | :|| |  || :
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db    202 SCFNGGTCVDGINSFTCLCPPGFTGSYCQ 230
          | |||:||||| |: ||| ||::| ||:
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT    9
ENTRY           A49175     #type fragment
TITLE           Motch B protein - mouse (fragment)
ALTERNATE_NAMES  Notch homolog
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                  14-Aug-1998
ACCESSIONS      A49175; PH1570; S32113
REFERENCE       A49175
   #authors     Lardelli, M.; Lendahl, U.
   #journal     Exp. Cell Res. (1993) 204:364-372
   #title       Motch A and Motch B--two mouse Notch homologues coexpressed
                  in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession   A49175
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues     1-1203 ##label LAR
      ##cross-references EMBL:X68279; NID:g287989; PID:g287990
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:126158)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermaland neural fate for cells in the early
             embryo.
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   560-591            #domain EGF homology #label EGF
SUMMARY         #length 1203  #checksum 910

  Query Match           9.9%;  Score 540;  DB 2;  Length 1203;
```

```
  Best Local Similarity  38.8%;  Pred. No. 6.30e-78;
  Matches   80;  Conservative   50;  Mismatches  57;  Indels  19;  Gaps   6;

Db    558 DECISKPCMNNGVCHNTQGS-YVCECPPGFSGMDCEEDINDCLANPCQNGGSC-VDHVNT 615
          | |:: ||:||::|  | :  | |:| ||| |: ||::|: | ::|| | ::| | :
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    616 FSCQCHPGFIGDKCQTDMNECLSEPCKNGGTC-------SD-------Y-VNSYTCTCPA 660
          |:| |: || || |: ::::|::  | ||| |       |:       : :||| | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    661 GFHGVHCENNIDECTE--SSCFNGGTCVDGINSFSCLCPVGFTGPFCLHDINECSSNPCL 718
          : | |||:::: ||   ::| | | |:    |:||:|: |||| |  :|::| :  |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    719 NAGTCVDGLGTYRCICPLGYTGKNCQ 744
          |:|:|||||: :| |:|  ||:|  |:
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT  10
ENTRY           A35844     #type complete
TITLE           Xotch protein - African clawed frog
ORGANISM        #formal_name Xenopus laevis #common_name African clawed frog
DATE            12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A35844
REFERENCE       A35844
  #authors      Coffman, C.; Harris, W.; Kintner, C.
  #journal      Science (1990) 249:1438-1441
  #title        Xotch, the Xenopus homolog of Drosophila notch.
  #cross-references MUID:90385285
  #accession    A35844
    ##status       preliminary; nucleic acid sequence not shown; not
                    compared with conceptual translation
    ##molecule_type mRNA
    ##residues     1-2524 ##label COF
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        transmembrane protein
FEATURE
  222-254           #domain EGF homology #label EGF\
  1924-1956         #domain ankyrin repeat homology #label AN1\
  1957-1989         #domain ankyrin repeat homology #label AN2\
  1991-2023         #domain ankyrin repeat homology #label AN3\
  2024-2056         #domain ankyrin repeat homology #label AN4\
  2057-2089         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2524  #molecular-weight 274931  #checksum 9441

  Query Match           9.9%;  Score 540;  DB 2;  Length 2524;
  Best Local Similarity  37.7%;  Pred. No. 6.30e-78;
  Matches   95;  Conservative   49;  Mismatches  81;  Indels  27;  Gaps  14;

Db    757 CESNPCMNGGTCKDMTGA-YICTCKAGFSGPNCQTNINECSSNPCLNHGTC-IDDVAGYK 814
          | ::||:| : |   ::  | | | :|| | :|:  |: | ::||||::|| : : : ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    815 CNCMLPYTGAICEAVLAPCAGSPCKNGGRCKE-----SED---PETF---S--CECPPGW 861
          | |   : |  ||  :  |  | | |||:| :     ||:   |::|   |  |:||
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    862 QGQTCEIDMNECVNR--PCRNGATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNG 919
          :|  ||  :: | ::  || | : |   |||| | | ||:|| ||| :||||    |:||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    920 GSCSDGINMFFCNCPAGFRGPKCE-EDI-N-ECA-SNPCKNGA-NCTDCV---NS--YTC 969
          ||| |||  : | | :|: |  ||   : : |   :::|  :|    :||   ||  :||
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    970 TCQPGFSGIHCE 981
           |: ||||  |:
Qy    423 KCHEGFSGPSCD 434


RESULT  11
ENTRY           I50719     #type complete
TITLE           C-Delta-1 - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  14-Aug-1998
ACCESSIONS      I50719
REFERENCE       I50719
  #authors      Henrique, D.; Adam, J.; Myat, A.; Chitnis, A.; Lewis, J.;
                  Ish-Horowicz, D.
  #journal      Nature (1995) 375:787-790
  #title        Expression of a Delta homologue in prospective neurons in the
                  chick.
  #cross-references MUID:95319507
  #accession    I50719
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues     1-728 ##label HEN
    ##cross-references EMBL:U26590; NID:g882411; PID:g882412
CLASSIFICATION  #superfamily EGF homology
FEATURE
  454-485           #domain EGF homology #label EGF
SUMMARY         #length 728  #molecular-weight 79861  #checksum 1765

  Query Match           9.8%;  Score 533;  DB 2;  Length 728;
  Best Local Similarity  38.7%;  Pred. No. 1.50e-76;
  Matches   79;  Conservative   40;  Mismatches  68;  Indels  17;  Gaps  12;

Db    302 HKPCKNGATCTNTGQGSYTCSCRPGYTGSSCEIEINECDANPCKNGGSCTDLENS-YSCT 360
          : |||| | |  |:   |||:| ||: |  || :|: | ::|| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    361 CPPGFYGKNCE--LSA-M-T-CADGP-CFNGGR-CTD---N--P-D-GGYSCRCPLGYSG 406
          |  || |  ||  :   : : | :|  | :  | |::   |  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    407 FNCEKKIDYCSS--SPCANGAQCVDLGNSYICQCQAGFTGRHCDDNVDDCASFPCVNGGT 464
           :|| |::||:   :|| | : |: :  || | | :|||| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    465 CQDGVNDYSCTCPPGYNGKNCSTP 488
          | ||:  | | | |||  |  |  |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIP 388


RESULT  12
ENTRY           I48324     #type complete
TITLE           DELTA-like 1 - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                  28-Feb-1997
ACCESSIONS      I48324
REFERENCE       I48324
  #authors      Bettenhausen, B.; de Angelis, M.H.; Simon, D.; Guenet, J.L.;
                  Gossler, A.
  #journal      Development (1995) 121:2407-2418
  #title        Transient and restricted expression during mouse
                  embryogenesis of Dll1, a murine gene closely related to
                  Drosophila Delta.
  #cross-references MUID:95401858
  #accession    I48324
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues     1-722 ##label RES
    ##cross-references EMBL:X80903; NID:g806569; PID:g806570
GENETICS
  #gene         Dll1
SUMMARY         #length 722  #molecular-weight 78448  #checksum 1452

  Query Match           9.7%;  Score 527;  DB 2;  Length 722;
  Best Local Similarity  35.6%;  Pred. No. 2.27e-75;
  Matches   89;  Conservative   54;  Mismatches  80;  Indels  27;  Gaps  20;
```

```
Db    294 HKPCRNGATCTNTGQGSYTCSCRPGYTGANCELEVDECAPSPCKNGASCTDLEDS-FSCT 352
          : ||:| | |  |:   |||:| ||: | :|| ::| |  ||| | |:|   : : |:|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    353 CPPGFYGKVCE--LSA-M-T-CADGP-CFNGGR-CSD---N--P-D-GGYTCHCPLGFSG 398
          |  || |  ||  :   : : | :|  | :  | ||:   |  : :  :| | ||: : |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    399 FNCEKKMDLCGS--SPCSNGAKCVDLGNSYLCRCQAGFSGRYCEDNVDDCASSPCANGGT 456
          :|| |:: |    :|| | :||: :  || | | :||:|  || |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    457 CRDSVNDFSCTCPPGYTGKNCS-APV-SR-CEHA-PCH-N--G-ATC--HQRGQRYMCEC 506
          | |::  : | | |||:|  |  :|:     : : :|: :  | : |   | :  : | |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKC 424

Db    507 AQGYGGPNCQ 516
          :|::||:|:
Qy    425 HEGFSGPSCD 434


RESULT  13
ENTRY           A35672     #type complete
TITLE           crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A35672
REFERENCE       A35672
  #authors      Tepass, U.; Theres, C.; Knust, E.
  #journal      Cell (1990) 61:787-799
  #title        crumbs encodes an EGF-like protein expressed on apical
                  membranes of Drosophila epithelial cells and required for
                  organization of epithelia.
  #cross-references MUID:90263104
  #accession    A35672
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-2139 ##label TEP
    ##cross-references GB:M33753
    ##note         the authors translated the codon GGC for residue 1928 as
                    Cys, and TAT for residue 2023 as Gln
GENETICS
  #gene         FlyBase:crb
    ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
  691-722           #domain EGF homology #label EGF
  MARY          #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           9.7%;  Score 529;  DB 2;  Length 2139;
  Best Local Similarity  30.5%;  Pred. No. 9.19e-76;
  Matches  100;  Conservative   80;  Mismatches 116;  Indels  32;  Gaps  25;

Db    747 FTCNC-IPGMRGRICDIDIDDCVGDPCLNGGQCIDQLG-GFRCDCSGTGYEGENCELNID 804
          |||:  :| : :  ||:    |:  || | : |    :  : |:|:  |: | :||  ||
Qy    166 FTCDSKVPTKLATKCDL----CLNSPCKNNAICETTSSRKYTCNCT-PGFYGVHCENQID 220

Db    805 ECLSNPCTNGAKC-LDRVKDYFCDCHNGYKGKNCEQDINECESNPCQYNGNCLERSNITL 863
          | ::|| | | | : :   : | |::|: |  || :|::| :: |:  |:|::   :
Qy    221 ACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDL--VR- 277

Db    864 YQMSRITDLPKVFSQPFSFENASGYECVCVPGIIGKNCEININECDS--NPCSKHGNCND 921
          :     :| | | |:| : | | | | |     ||:|| ::: |    ||| ::|:|
Qy    278 F--CS-EEL-KNF-QSFQI-N-S-YRCDCPMEYEGKHCEDKLEYCTKKLNPCENNGKCIP 329

Db    922 GIGTYTCECEPGFEGTHCEINIDECDRYNPCQRG-TCYDQIDDYDCDCDANYGGKNCSV- 979
            |:|:| | ||| | :|| |||:| :   || | :| | |  ||| | : |:|  | :
Qy    330 INGSYSCMCSPGFTGNNCETNIDDC-KNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIP 388

Db    980 -LL-KGCDQ-NPCLNGGAC-LP-YLINEVTHLYTCTCENGFQGDKCEKTTTLS-MVATSL 1033
           :: :  :  ::|  : ||     : :: :  :|| | :|| |  |::  ::: : : :
Qy    389 PMMDMEYQKTDACQQS-ACGQGECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAY 447

Db   1034 ISVTTEREEGYDINLQFRTTLPNGVLAF 1061
          :::    :|  | : :|||   |:| :
Qy    448 LALDPLASDG-TITMTLRTTSKIGILLY 474


RESULT  14
ENTRY           A49128     #type complete
TITLE           cell-fate determining gene Notch2 protein - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A49128
REFERENCE       A49128
  #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
  #journal      Development (1992) 116:931-941
  #title        Notch2: a second mammalian Notch gene.
  #cross-references MUID:93202015
  #accession    A49128
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type mRNA
    ##residues     1-2471 ##label WEI
    ##experimental_source Schwann cell
    ##note         sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
  1029-1060         #domain EGF homology #label EGF\
  1876-1908         #domain ankyrin repeat homology #label AN1\
  1909-1941         #domain ankyrin repeat homology #label AN2\
  1943-1975         #domain ankyrin repeat homology #label AN3\
  1976-2008         #domain ankyrin repeat homology #label AN4\
  2009-2041         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           9.7%;  Score 527;  DB 2;  Length 2471;
  Best Local Similarity  39.3%;  Pred. No. 2.27e-75;
  Matches   81;  Conservative   46;  Mismatches 60;  Indels  19;  Gaps   6;

Db    875 DECVSKPCMNNGICHNTQGS-YMCECPPGFSGMDCEEDINDCLANPCQNGGSC-VDKVNT 932
          | |:: ||:||:||  | :  | |:| ||| |: ||::|: | ::|| | ::| |
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    933 FSCLCLPGFVGDKCQTDMNECLSEPCKNGGTC-------SD-------Y-VNSYTCTCPA 977
          |:| |  || || |: ::::|::  | ||| |       |:       : :||| | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    978 GFHGVHCENNIDECTE--SSCFNGGTCVDGINSFSCLCPVGFTGPFCLHDINECSSNPCL 1035
          : | |||:::: ||   ::| | | |:    |:||:|: |||| |  :|::| :  |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db   1036 NSGTCVDGLGTYRCTCPLGYTGKNCQ 1061
          |:|:|||||: :| | |  ||:|  |:
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCE 386


RESULT  15
ENTRY           A31246     #type complete
TITLE           neurogenic protein Delta precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A31246
REFERENCE       A31246
  #authors      Kopczynski, C.C.; Alton, A.K.; Fechtel, K.; Kooh, P.J.;
                  Muskavitch, M.A.T.
  #journal      Genes Dev. (1988) 2:1723-1735
  #title        Delta, a Drosophila neurogenic gene, is transcriptionally
                  complex and encodes a protein related to blood coagulation
                  factors and epidermal growth factor of vertebrates.
```

```
   #cross-references MUID:89196890
   #accession    A31246
      ##molecule_type mRNA
      ##residues      1-832 ##label KOP
      ##cross-references GB:Y00222
GENETICS
   #gene         FlyBase:Dl
      ##cross-references FlyBase:FBgn0000463
CLASSIFICATION   #superfamily EGF homology
FEATURE
   457-488            #domain EGF homology #label EGF
SUMMARY          #length 832  #molecular-weight 88943  #checksum 636

  Query Match            9.6%;  Score 520;  DB 2;  Length 832;
  Best Local Similarity  37.1%;  Pred. No. 5.38e-74;
  Matches    95;  Conservative   55;  Mismatches  77;  Indels  29;  Gaps  22;

Db    295 CTNHRPCKNGGTCFNTGEGLYTCKCAPGYSGDDCENEIYSCDADVNPCQNGGTCIDEPHT 354
          | |  |||| : |  |:   |||:|:||: |  |||:| :| :  :|| | :||   :::
Qy    183 CLN-SPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYG--SPCLNNATC-KVAQA 238

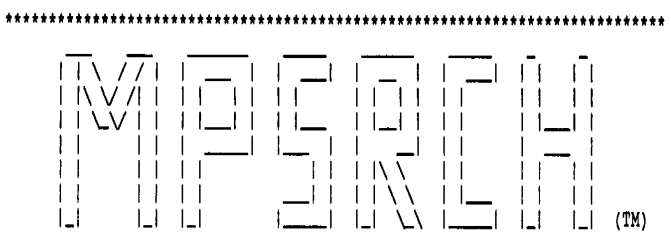
Db    355 KTGYKCHCRNGWSGKMCEEKVLTCSDKPCHQG-ICRN-VR--PG-LGS-KG-Q--GYQCE 405
            ::| | :|  |  || ::  | :  |  |  | : ||  :  | :  : |  :|:|:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db    406 CPIGYSGPNCDLQLDNCSP--NPCINGGSCQP-SGK--CICPSGFSGTRCETNIDDCLGH 460
          ||: | | :|:  |: |:   ||| | | | | :|   |:|::||:|  ||||||||
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db    461 QCENGGTCIDMVNQYRCQCVPGFHGTHC--SSKVDL-CL-IRPC--AN-G-GTCL-NLNN 511
          :|:|||:|:| :  | | | ||: |  |  ::::|:     :|  :  | | |: : |:
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db    512 -DYQCTCRAGFTGKDC 526
           |: | |: ||:|  |
Qy    418 SDFTCKCHEGFSGPSC 433


Search completed: Fri May 28 09:11:33 1999
Job time : 79 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:11:51 1999; MasPar time 21.65 Seconds
959.467 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-9
Description: (1-735) from US09191647.pep
Perfect Score: 5438
Sequence: 1 SNKNLTSFPSRIPFDTTELY.....TVHIIRQCQCEPTKSVLSEK 735

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 51.090; Variance 86.159; scale 0.593

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1838	33.8	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	0.00e+00
2	588	10.8	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	4.02e-104
3	559	10.3	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	2.70e-97
4	557	10.2	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	7.95e-97
5	548	10.1	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	1.03e-94
6	542	10.0	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	2.61e-93
7	527	9.7	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	8.40e-90
8	529	9.7	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	2.87e-90
9	529	9.7	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	2.87e-90
10	520	9.6	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	3.61e-88
11	520	9.6	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	3.61e-88
12	518	9.5	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	1.06e-87
13	508	9.3	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	2.25e-85
14	505	9.3	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	1.12e-84
15	506	9.3	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	6.56e-85
16	481	8.8	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	4.10e-79
17	434	8.0	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	2.59e-68
18	417	7.7	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	1.93e-64
19	412	7.6	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	2.63e-63
20	401	7.4	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	8.14e-61
21	391	7.2	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	1.47e-58
22	318	5.8	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	2.24e-42
23	304	5.6	3358	1	PGCV_MOUSE	VERSICAN CORE PROTEIN	2.44e-39
24	307	5.6	3396	1	PGCV_HUMAN	VERSICAN CORE PROTEIN	5.47e-40
25	300	5.5	862	1	PGCV_MACNE	VERSICAN CORE PROTEIN	1.78e-38
26	292	5.4	3562	1	PGCV_CHICK	VERSICAN CORE PROTEIN	9.29e-37
27	285	5.2	886	1	EMR1_HUMAN	CELL SURFACE GLYCOPROT	2.91e-35
28	276	5.1	1257	1	PGCN_RAT	NEUROCAN CORE PROTEIN	2.38e-33
29	275	5.1	1394	1	TGFB_HUMAN	LATENT TRANSFORMING GR	3.88e-33
30	278	5.1	1959	1	AGRI_RAT	AGRIN PRECURSOR.	8.97e-34
31	272	5.0	1245	1	NIDO_MOUSE	NIDOGEN PRECURSOR (ENT	1.67e-32
32	274	5.0	1268	1	PGCN_MOUSE	NEUROCAN CORE PROTEIN	6.31e-33
33	267	4.9	1712	1	TGFB_RAT	LATENT TRANSFORMING GR	1.89e-31
34	269	4.9	2911	1	FBN2_HUMAN	FIBRILLIN 2 PRECURSOR.	7.16e-32
35	259	4.8	816	1	NEL_CHICK	NEL PROTEIN PRECURSOR	8.95e-30
36	263	4.8	2871	1	FBN1_BOVIN	FIBRILLIN 1 PRECURSOR	1.30e-30
37	262	4.8	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	2.11e-30
38	261	4.8	2871	1	FBN1_HUMAN	FIBRILLIN 1 PRECURSOR.	3.42e-30
39	254	4.7	816	1	NEL_HUMAN	NEL PROTEIN PRECURSOR	9.85e-29
40	255	4.7	2907	1	FBN2_MOUSE	FIBRILLIN 2 PRECURSOR.	6.10e-29
41	243	4.5	816	1	NEL_MOUSE	NEL PROTEIN PRECURSOR	1.85e-26
42	246	4.5	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	4.47e-27
43	239	4.4	816	1	NEL_RAT	NEL PROTEIN PRECURSOR	1.23e-25
44	240	4.4	1247	1	NIDO_HUMAN	NIDOGEN PRECURSOR (ENT	7.65e-26
45	234	4.3	1376	1	NID2_HUMAN	NIDOGEN-2 PRECURSOR (N	1.29e-24

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE  1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          33.8%;  Score 1838;  DB 1;  Length 1480;
  Best Local Similarity  41.8%;  Pred. No. 0.00e+00;
  Matches  257;  Conservative  143;  Mismatches 194;  Indels  21;  Gaps  15;

Db    729 SRNQLKEIPRGIPAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFAN 788
          | : |  :|  || :|:|||||::| |::|    : :| |||:||||:|::  | | ||:|
Qy      1 SNKNLTSFPSRIPFDTTELYLDANYINEIPAHDLNRLYSLTKLDLSHNRLISLENNTFSN 60

Db    789 LTKLSTLIISYNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSN 848
          ||:|||||||||||:|||  |::||| ||::||||| || ||:::| :| |:||||:|||
Qy     61 LTRLSTLIISYNKLRCLQPLAFNGLNALRILSLHGNDISFLPQSAFSNLTSITHIAVGSN 120

Db    849 PLYCDCGLKWFSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKC 908
          :||||| : ||| |||  ::|:|||||  |: : : |:|:: :  | | ::|   : :||
Qy    121 SLYCDCNMAWFSKWIKSKFIEAGIARCEYPNTVSNQLLLTAQPYQFTCDSKVPTKLATKC 180

Db    909 NACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEPMIDACYGNPCRNNATCTVLEEGR 968
          : |:: || | | | : : | | | | ||: | |||  ||||||:|| ||||| | : ||
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    969 FSCQCAPGYTGARCETNIDDCL-G--EI--KCQNNAT-C---I-D--G--VESYKCECQP 1014
          |:| |  |: |  || |||||: :  |   || :    |   : :  :  ::||:|:|
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db   1015 GFSGEFCDTKIQFCSPEFNPCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQ 1074
           : |  |: |:::|:  :||| | :||:     ||| | :|| | ||  ||||| |  ||
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db   1075 NGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSD 1134
          |||:||||| |:| |  |:| |||  |: | |  | :||  | :| |  ::  :||
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECV-ASQNSSD 419

Db   1135 YLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQ 1194
          : |:|| |::|  |:   |::| :  ::: |:|| : ::: :|: : ::   |||:| |:
Qy    420 FTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDPL-A-SDGTITMTLRTTSKIGILLYYGD 477

Db   1195 DAHLAVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVEL-LAIKKNFTLRVDRGLA 1253
          |  :: ||::||::: | :|| | | |||  :| ||  | : :  : :| | |::|:
Qy    478 DHFVSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCF-LQIDKNPV 536

Db   1254 RSIINEGSNDYLKLT-TPM-FLGGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVD 1311
          : : | | :| |      | ::||||:: :|:| :  :::|  |:|||:  : ||   ::
Qy    537 QIVENSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEVPIN 596

Db   1312 FGNAQRQQKITPGCA 1326
          :  |    :   :|:
Qy    597 LQQALENVNTEQSCS 611


RESULT    2
```

```
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1064       FIBROPELLIN I.
FT   DOMAIN      20     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480    516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     518    554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     556    592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     594    630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632    668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     670    706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     708    744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     746    782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     784    820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     822    858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     860    896       EGF-LIKE 20.
FT   DOMAIN     898    934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     936   1064       AVIDIN-LIKE.
FT   DISULFID    23     34       BY SIMILARITY.
FT   DISULFID    28     43       BY SIMILARITY.
FT   DISULFID    45     54       BY SIMILARITY.
FT   DISULFID   180    191       BY SIMILARITY.
FT   DISULFID   185    200       BY SIMILARITY.
FT   DISULFID   202    211       BY SIMILARITY.
FT   DISULFID   218    229       BY SIMILARITY.
FT   DISULFID   223    238       BY SIMILARITY.
FT   DISULFID   240    249       BY SIMILARITY.
FT   DISULFID   256    267       BY SIMILARITY.
FT   DISULFID   261    276       BY SIMILARITY.
FT   DISULFID   278    287       BY SIMILARITY.
FT   DISULFID   294    305       BY SIMILARITY.
FT   DISULFID   299    314       BY SIMILARITY.
FT   DISULFID   316    325       BY SIMILARITY.
FT   DISULFID   332    343       BY SIMILARITY.
FT   DISULFID   337    352       BY SIMILARITY.
FT   DISULFID   354    363       BY SIMILARITY.
FT   DISULFID   370    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   430    439       BY SIMILARITY.
FT   DISULFID   446    457       BY SIMILARITY.
FT   DISULFID   451    466       BY SIMILARITY.
FT   DISULFID   468    477       BY SIMILARITY.
FT   DISULFID   484    495       BY SIMILARITY.
FT   DISULFID   489    504       BY SIMILARITY.
FT   DISULFID   506    515       BY SIMILARITY.
FT   DISULFID   522    533       BY SIMILARITY.
FT   DISULFID   527    542       BY SIMILARITY.
FT   DISULFID   544    553       BY SIMILARITY.
FT   DISULFID   560    571       BY SIMILARITY.
FT   DISULFID   565    580       BY SIMILARITY.
FT   DISULFID   582    591       BY SIMILARITY.
```

```
FT   DISULFID    598    609       BY SIMILARITY.
FT   DISULFID    603    618       BY SIMILARITY.
FT   DISULFID    620    629       BY SIMILARITY.
FT   DISULFID    636    647       BY SIMILARITY.
FT   DISULFID    641    656       BY SIMILARITY.
FT   DISULFID    658    667       BY SIMILARITY.
FT   DISULFID    674    685       BY SIMILARITY.
FT   DISULFID    679    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    712    723       BY SIMILARITY.
FT   DISULFID    717    732       BY SIMILARITY.
FT   DISULFID    734    743       BY SIMILARITY.
FT   DISULFID    750    761       BY SIMILARITY.
FT   DISULFID    755    770       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    788    799       BY SIMILARITY.
FT   DISULFID    793    808       BY SIMILARITY.
FT   DISULFID    810    819       BY SIMILARITY.
FT   DISULFID    826    837       BY SIMILARITY.
FT   DISULFID    831    846       BY SIMILARITY.
FT   DISULFID    848    857       BY SIMILARITY.
FT   DISULFID    864    875       BY SIMILARITY.
FT   DISULFID    869    884       BY SIMILARITY.
FT   DISULFID    886    895       BY SIMILARITY.
FT   DISULFID    902    913       BY SIMILARITY.
FT   DISULFID    907    922       BY SIMILARITY.
FT   DISULFID    924    933       BY SIMILARITY.
FT   VARSPLIC    477    780       MISSING (IN FORM IB).
FT   CARBOHYD     30     30       POTENTIAL.
FT   CARBOHYD    136    136       POTENTIAL.
FT   CARBOHYD    851    851       POTENTIAL.
FT   CONFLICT    279    279       L -> S (IN REF. 2).
SQ   SEQUENCE   1064 AA;  112072 MW;  FBD10D48 CRC32;

  Query Match          10.8%;  Score 588;  DB 1;  Length 1064;
  Best Local Similarity  39.7%;  Pred. No. 4.02e-104;
  Matches  102;  Conservative  47;  Mismatches  79;  Indels  29;  Gaps  17;

Db    216 DECASDPCQNGGAC-VDGVNGYVCNCVPGFDGDECENNINECASSPCLNGGIC-VDGVNM 273
          | | : || | : |   :   | ||| ||| |  ||| |: | :||||| : | |
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    274 FECTCLAGFTGVRCEVNIDECASAPCQNGG-----I--C---I-D--G--INGYTCSCPL 318
          |:| |  || |  || |||:| :: |:|||     :  |   : :  :  ||:| | ||:
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    319 GFSGDNCENNDDECSS-I-PCLNGGTCVDLVNAYMCVCAPGWTGPTCADNIDECASAPCQ 376
          : | :||:: : |:  : || | | |: :  :| |:|:|| ||  |  |||:| :  ||
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db    377 NGGVCIDGVNGYMCDCQPGYTGTHCETD--ID-ECARP-PCQ-N--G-GDCVDGVNG--Y 426
          ||| |:||: :| | |:|||:|  ||    :| |  :  :|| :  | |:|| : |:  :
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDF 420

Db    427 VCICAPGFDGLNCENNI 443
          | |  || | :|:  :
Qy    421 TCKCHEGFSGPSCDRQM 437


RESULT    3
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    570       FIBROPELLIN C.
FT   DOMAIN       18     55       EGF-LIKE 1.
FT   DOMAIN       62    175       CUB.
FT   DOMAIN      176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      366    402       EGF-LIKE 7.
FT   DOMAIN      404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      442    570       AVIDIN-LIKE.
FT   DISULFID     23     34       BY SIMILARITY.
FT   DISULFID     28     43       BY SIMILARITY.
FT   DISULFID     45     54       BY SIMILARITY.
FT   DISULFID    180    191       BY SIMILARITY.
FT   DISULFID    185    200       BY SIMILARITY.
FT   DISULFID    202    211       BY SIMILARITY.
FT   DISULFID    218    229       BY SIMILARITY.
FT   DISULFID    223    238       BY SIMILARITY.
FT   DISULFID    240    249       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    261    276       BY SIMILARITY.
FT   DISULFID    278    287       BY SIMILARITY.
FT   DISULFID    294    305       BY SIMILARITY.
FT   DISULFID    299    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
```

```
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   CARBOHYD     30     30       POTENTIAL.
FT   CARBOHYD    136    136       POTENTIAL.
FT   CARBOHYD    357    357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match          10.3%;  Score 559;  DB 1;  Length 570;
  Best Local Similarity  42.2%;  Pred. No. 2.70e-97;
  Matches   86;  Conservative   35;  Mismatches  64;  Indels  19;  Gaps   8;

Db    256 CASIPCLNGGIC-VDGINQFACTCLPGYTGILCETDINECASSPCQNGGSCTDA-VNRYT 313
          | : || | :||   :   ::| | ||: |: || :|: | :||| | ::|  |   |:
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSC--L------D-----G--VDGYVCQCLPNY 358
          | |  || | || ||::| :| | ||| |  |      :     : :::| |:|  :|
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    359 TGTHCEISLDACAS-L-PCQNGGVCTNVGGDYVCECLPGYTGINCEIDINECASLPCQNG 416
          | |||  |: |:  | ||:| | |  : | | | | ||:|| ||| :|::| :: ||||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    417 GECINGIAMYICQCRQGYAGVNCE 440
          | |::||  | | || ||||  ||
Qy    363 GSCVDGILSYDCLCRPGYAGQYCE 386


RESULT   4
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     44       POTENTIAL.
FT   CHAIN       45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746   1766       POTENTIAL.
FT   DOMAIN    1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58     95       EGF-LIKE 1.
FT   DOMAIN      96    136       EGF-LIKE 2.
FT   DOMAIN     139    176       EGF-LIKE 3.
FT   DOMAIN     177    215       EGF-LIKE 4.
FT   DOMAIN     217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255    291       EGF-LIKE 6.
FT   DOMAIN     293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     409    447       EGF-LIKE 10.
FT   DOMAIN     449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      867    905       EGF-LIKE 22.
FT   DOMAIN      907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984   1020       EGF-LIKE 25.
FT   DOMAIN     1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1060   1096       EGF-LIKE 27.
FT   DOMAIN     1098   1134       EGF-LIKE 28.
FT   DOMAIN     1136   1181       EGF-LIKE 29.
FT   DOMAIN     1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1297   1335       EGF-LIKE 33.
FT   DOMAIN     1337   1373       EGF-LIKE 34.
FT   DOMAIN     1375   1412       EGF-LIKE 35.
FT   DOMAIN     1415   1451       EGF-LIKE 36.
FT   DOMAIN     1475   1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1475   1513       LIN/NOTCH 1.
FT   REPEAT     1514   1553       LIN/NOTCH 2.
FT   REPEAT     1554   1593       LIN/NOTCH 3.
FT   DOMAIN     1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN     2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID     62     73       BY SIMILARITY.
FT   DISULFID     67     83       BY SIMILARITY.
FT   DISULFID     85     94       BY SIMILARITY.
FT   DISULFID    100    111       BY SIMILARITY.
FT   DISULFID    105    124       BY SIMILARITY.
FT   DISULFID    126    135       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    181    192       BY SIMILARITY.
FT   DISULFID    186    203       BY SIMILARITY.
FT   DISULFID    205    214       BY SIMILARITY.
FT   DISULFID    221    232       BY SIMILARITY.
FT   DISULFID    226    241       BY SIMILARITY.
FT   DISULFID    243    252       BY SIMILARITY.
FT   DISULFID    259    270       BY SIMILARITY.
FT   DISULFID    264    279       BY SIMILARITY.
FT   DISULFID    281    290       BY SIMILARITY.
FT   DISULFID    297    308       BY SIMILARITY.
FT   DISULFID    302    317       BY SIMILARITY.
FT   DISULFID    319    328       BY SIMILARITY.
FT   DISULFID    335    349       BY SIMILARITY.
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    396       BY SIMILARITY.
FT   DISULFID    398    407       BY SIMILARITY.
FT   DISULFID    413    424       BY SIMILARITY.
FT   DISULFID    418    435       BY SIMILARITY.
FT   DISULFID    437    446       BY SIMILARITY.
FT   DISULFID    453    465       BY SIMILARITY.
FT   DISULFID    459    474       BY SIMILARITY.
FT   DISULFID    476    485       BY SIMILARITY.
FT   DISULFID    492    503       BY SIMILARITY.
FT   DISULFID    497    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    579       BY SIMILARITY.
FT   DISULFID    573    588       BY SIMILARITY.
FT   DISULFID    590    599       BY SIMILARITY.
FT   DISULFID    606    616       BY SIMILARITY.
FT   DISULFID    611    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    654       BY SIMILARITY.
FT   DISULFID    648    663       BY SIMILARITY.
FT   DISULFID    665    674       BY SIMILARITY.
FT   DISULFID    681    692       BY SIMILARITY.
FT   DISULFID    686    701       BY SIMILARITY.
FT   DISULFID    703    712       BY SIMILARITY.
FT   DISULFID    719    730       BY SIMILARITY.
FT   DISULFID    724    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          10.2%;  Score 557;  DB 1;  Length 2703;
  Best Local Similarity  38.1%;  Pred. No. 7.95e-97;
  Matches   96;  Conservative   49;  Mismatches  80;  Indels  27;  Gaps  17;

Db    681 CHSNPCNNGATC-IDGINSYKCQCVPGFTGQHCEKNVDECISSPCANNGVC-IDQVNGYK 738
          | ::||:| | |   :   | | | ||| | |||: :| | :|| ||: | : |   ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    739 CECPRGFYDAHCLSDVDECA-S---NP--CVNEGR-CEDGI-N-E-F-I----CHCPPGY 783
          | | :||    | ::|:| |   |   ||: | | : : | : | |    | ||  |
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    784 TGKRCELDIDECSS--NPCQHGGTCYDKLNAFSCQCMPGYTGQKCETNIDDCVTNPCGNG 841
           ||:||  :: |:   |||:: | |     ::|| | ||:|| :|||||||    | ||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    842 GTCIDKVNGYKCVCKVPFTGRDCE--SKMD-PCA-SNRCK-N---EAKCTPSSNFLDFSC 893
          |:|:| : :| |:|:  ::|: ||  ::||     :: | :   :: | :| | ||:|
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    894 TCKLGYTGRYCD 905
           |  |::|  ||
Qy    423 KCHEGFSGPSCD 434


RESULT   5
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
```

```
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21   1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725   1747       POTENTIAL.
FT   DOMAIN    1748   2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     57       EGF-LIKE 1.
FT   DOMAIN      58     98       EGF-LIKE 2.
FT   DOMAIN     101    138       EGF-LIKE 3.
FT   DOMAIN     139    175       EGF-LIKE 4.
FT   DOMAIN     177    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     827    865       EGF-LIKE 22.
FT   DOMAIN     867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     981   1017       EGF-LIKE 26.
FT   DOMAIN    1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1057   1093       EGF-LIKE 28.
FT   DOMAIN    1095   1141       EGF-LIKE 29.
FT   DOMAIN    1143   1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1181   1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1219   1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1265   1303       EGF-LIKE 33.
FT   DOMAIN    1305   1344       EGF-LIKE 34.
FT   DOMAIN    1346   1382       EGF-LIKE 35.
FT   DOMAIN    1385   1423       EGF-LIKE 36.
FT   DOMAIN    1446   1561       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446   1486       LIN/NOTCH 1.
FT   REPEAT    1487   1520       LIN/NOTCH 2.
FT   REPEAT    1521   1561       LIN/NOTCH 3.
FT   DOMAIN    1861   2074       6 X ANK MOTIF REPEATS.
FT   REPEAT    1861   1891       ANK MOTIF 1.
FT   REPEAT    1892   1940       ANK MOTIF 1.
FT   REPEAT    1941   1974       ANK MOTIF 1.
FT   REPEAT    1975   2007       ANK MOTIF 1.
FT   REPEAT    2008   2040       ANK MOTIF 1.
FT   REPEAT    2041   2074       ANK MOTIF 1.
FT   DOMAIN    2265   2276       POLY-GLN (OPA-REPEAT).
FT   DISULFID    25     35       BY SIMILARITY.
FT   DISULFID    29     45       BY SIMILARITY.
FT   DISULFID    47     56       BY SIMILARITY.
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     86       BY SIMILARITY.
FT   DISULFID    88     97       BY SIMILARITY.
FT   DISULFID   105    116       BY SIMILARITY.
FT   DISULFID   110    126       BY SIMILARITY.
FT   DISULFID   128    137       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    163       BY SIMILARITY.
FT   DISULFID   165    174       BY SIMILARITY.
FT   DISULFID   181    194       BY SIMILARITY.
FT   DISULFID   188    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    242       BY SIMILARITY.
FT   DISULFID   244    253       BY SIMILARITY.
FT   DISULFID   260    271       BY SIMILARITY.
FT   DISULFID   265    280       BY SIMILARITY.
FT   DISULFID   282    291       BY SIMILARITY.
FT   DISULFID   298    311       BY SIMILARITY.
FT   DISULFID   305    320       BY SIMILARITY.
FT   DISULFID   322    331       BY SIMILARITY.
FT   DISULFID   338    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   375    386       BY SIMILARITY.
FT   DISULFID   380    397       BY SIMILARITY.
FT   DISULFID   399    408       BY SIMILARITY.
FT   DISULFID   415    428       BY SIMILARITY.
FT   DISULFID   422    437       BY SIMILARITY.
FT   DISULFID   439    448       BY SIMILARITY.
FT   DISULFID   455    466       BY SIMILARITY.
FT   DISULFID   460    475       BY SIMILARITY.
FT   DISULFID   477    486       BY SIMILARITY.
FT   DISULFID   493    503       BY SIMILARITY.
FT   DISULFID   498    512       BY SIMILARITY.
FT   DISULFID   514    523       BY SIMILARITY.
FT   DISULFID   530    541       BY SIMILARITY.
FT   DISULFID   535    550       BY SIMILARITY.
FT   DISULFID   552    561       BY SIMILARITY.
FT   DISULFID   568    578       BY SIMILARITY.
FT   DISULFID   573    587       BY SIMILARITY.
FT   DISULFID   589    598       BY SIMILARITY.
FT   DISULFID   605    616       BY SIMILARITY.
FT   DISULFID   610    625       BY SIMILARITY.
FT   DISULFID   627    636       BY SIMILARITY.
FT   DISULFID   643    653       BY SIMILARITY.
FT   DISULFID   648    662       BY SIMILARITY.
FT   DISULFID   664    673       BY SIMILARITY.
FT   DISULFID   680    691       BY SIMILARITY.
FT   DISULFID   685    700       BY SIMILARITY.
FT   DISULFID   702    711       BY SIMILARITY.
FT   DISULFID   718    728       BY SIMILARITY.
FT   DISULFID   723    737       BY SIMILARITY.
FT   DISULFID   739    748       BY SIMILARITY.
FT   DISULFID   755    766       BY SIMILARITY.
FT   DISULFID   760    775       BY SIMILARITY.
FT   DISULFID   777    786       BY SIMILARITY.
FT   DISULFID   793    804       BY SIMILARITY.
FT   DISULFID   798    813       BY SIMILARITY.
FT   DISULFID   815    824       BY SIMILARITY.
FT   DISULFID   831    842       BY SIMILARITY.
FT   DISULFID   836    853       BY SIMILARITY.
FT   DISULFID   855    864       BY SIMILARITY.
FT   DISULFID   871    882       BY SIMILARITY.
FT   DISULFID   876    891       BY SIMILARITY.
FT   DISULFID   893    902       BY SIMILARITY.
FT   DISULFID   909    920       BY SIMILARITY.
FT   DISULFID   914    929       BY SIMILARITY.
FT   DISULFID   931    940       BY SIMILARITY.
```

```
FT   DISULFID   947   958       BY SIMILARITY.
FT   DISULFID   952   967       BY SIMILARITY.
FT   DISULFID   969   978       BY SIMILARITY.
FT   DISULFID  1023  1034       BY SIMILARITY.
FT   DISULFID  1028  1043       BY SIMILARITY.
FT   DISULFID  1045  1054       BY SIMILARITY.
FT   DISULFID  1061  1072       BY SIMILARITY.
FT   DISULFID  1066  1081       BY SIMILARITY.
FT   DISULFID  1083  1092       BY SIMILARITY.
FT   DISULFID  1099  1120       BY SIMILARITY.
FT   DISULFID  1114  1129       BY SIMILARITY.
FT   DISULFID  1131  1140       BY SIMILARITY.
FT   DISULFID  1147  1158       BY SIMILARITY.
FT   DISULFID  1152  1167       BY SIMILARITY.
FT   DISULFID  1169  1178       BY SIMILARITY.
FT   DISULFID  1185  1196       BY SIMILARITY.
FT   DISULFID  1190  1205       BY SIMILARITY.
FT   DISULFID  1207  1216       BY SIMILARITY.
FT   DISULFID  1223  1242       BY SIMILARITY.
FT   DISULFID  1236  1251       BY SIMILARITY.
FT   DISULFID  1253  1262       BY SIMILARITY.

Note: remainder of annotations omitted.

  Query Match          10.1%;  Score 548;  DB 1;  Length 2437;
  Best Local Similarity  39.6%;  Pred. No. 1.03e-94;
  Matches  105;  Conservative  56;  Mismatches  70;  Indels  34;  Gaps  24;

Db    528 DECASTPCKNGAKC-TDGPNKYTCECTPGFSGIHCELDINECASSPC-HYGVCRDGVAS- 584
          | | ::|||| | | | :: ||||:||||| |:||| :|: | :||| : : |: : |:
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db    585 FTCDCRPGYTGRLCETNINECLSQPCRNGGTCQD--R----E--N--AY-I----CTCPK 629
          | | |  |: |  || ||::|::  | ||| | |  |    | |  :: |    | ||:
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db    630 GTTGVNCEINIDDC-KR-KPCDY-GKCIDKING-YECVCEPGYSGSMCNINIDDCALNPC 685
            | :|| ::: | |: :||:  ||||  ||| | |:| ||::|: |: |||||    |
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIP-INGSYSCMCSPGFTGNNCETNIDDCKNVEC 359

Db    686 HNGGTCIDGVNSFTCLC-PD--G-FRDATCL-S-Q-HN-E-CSSNPCIHGSCLD-QINS- 734
          :|||:|:||: |: ||| |   | : :   :   : :: : |  ::| :| |:  | :|
Qy    360 QNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSD 419

Db    735 YRCVCEAGWMGRNCDININECLSNP 759
          : | |  |  | :||  ::  : ||
Qy    420 FTCKCHEGFSGPSCDRQMSVGFKNP 444


RESULT    6
    NOTC_XENLA     STANDARD;      PRT;  2524 AA.
    P21783;
DT  01-MAY-1991 (REL. 18, CREATED)
DT  01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN  XOTCH.
OS  XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC  MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 90385285.
RA  COFFMAN C., HARRIS W., KINTNER C.;
RT  "Xotch, the Xenopus homolog of Drosophila notch.";
RL  SCIENCE 249:1438-1441(1990).
RN  [2]
RP  REVISIONS TO 1759-1782.
RA  KINTNER C.;
RL  SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC  -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC  -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC  -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC  -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC  -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC  -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; M33874; G1364263; -.
DR  PIR; A35844; A35844.
DR  PROSITE; PS00010; ASX_HYDROXYL; 23.
DR  PROSITE; PS00022; EGF_1; 34.
DR  PROSITE; PS01186; EGF_2; 29.
DR  PROSITE; PS01187; EGF_CA; 21.
DR  PFAM; PF00008; EGF; 36.
DR  PFAM; PF00023; ank; 6.
DR  PFAM; PF00066; notch; 3.
DR  HSSP; P00740; 1IXA.
KW  DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW  TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT  SIGNAL       1    19       POTENTIAL.
FT  CHAIN       20  2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT  DOMAIN      20  1728       EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM  1729  1750       POTENTIAL.
FT  DOMAIN    1751  2524       CYTOPLASMIC (POTENTIAL).
FT  DOMAIN      20    57       EGF-LIKE 1.
FT  DOMAIN      58    99       EGF-LIKE 2.
FT  DOMAIN     102   140       EGF-LIKE 3.
FT  DOMAIN     141   177       EGF-LIKE 4.
FT  DOMAIN     179   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     217   254       EGF-LIKE 6.
FT  DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     371   409       EGF-LIKE 10.
FT  DOMAIN     411   449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     451   487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     489   525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     527   563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     565   600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     602   638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     640   675       EGF-LIKE 17.
FT  DOMAIN     677   713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     715   750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     752   788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     790   826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     828   866       EGF-LIKE 22.
FT  DOMAIN     868   904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     906   942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     944   980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN     982  1018       EGF-LIKE 26.
FT  DOMAIN    1020  1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN    1058  1094       EGF-LIKE 28.
FT  DOMAIN    1096  1142       EGF-LIKE 29.
FT  DOMAIN    1144  1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN    1182  1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN    1220  1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN    1266  1304       EGF-LIKE 33.
FT  DOMAIN    1306  1346       EGF-LIKE 34.
FT  DOMAIN    1347  1383       EGF-LIKE 35.
FT  DOMAIN    1386  1424       EGF-LIKE 36.
FT  DOMAIN    1441  1560       3 X LIN/NOTCH REPEATS.
FT  REPEAT    1441  1478       LIN/NOTCH 1.
FT  REPEAT    1479  1520       LIN/NOTCH 2.
FT  REPEAT    1521  1560       LIN/NOTCH 3.
FT  DOMAIN    1871  2083       6 X ANK MOTIF REPEATS.
FT  DISULFID    22    35       BY SIMILARITY.
FT  DISULFID    29    45       BY SIMILARITY.
FT  DISULFID    47    56       BY SIMILARITY.
```

```
FT   DISULFID    62     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    128       BY SIMILARITY.
FT   DISULFID   130    139       BY SIMILARITY.
FT   DISULFID   145    156       BY SIMILARITY.
FT   DISULFID   150    165       BY SIMILARITY.
FT   DISULFID   167    176       BY SIMILARITY.
FT   DISULFID   183    194       BY SIMILARITY.
FT   DISULFID   188    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    242       BY SIMILARITY.
FT   DISULFID   244    253       BY SIMILARITY.
FT   DISULFID   260    271       BY SIMILARITY.
FT   DISULFID   265    280       BY SIMILARITY.
FT   DISULFID   282    291       BY SIMILARITY.
FT   DISULFID   298    311       BY SIMILARITY.
FT   DISULFID   305    320       BY SIMILARITY.
FT   DISULFID   322    331       BY SIMILARITY.
FT   DISULFID   338    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   375    386       BY SIMILARITY.
FT   DISULFID   380    397       BY SIMILARITY.
FT   DISULFID   399    408       BY SIMILARITY.
FT   DISULFID   415    428       BY SIMILARITY.
FT   DISULFID   422    437       BY SIMILARITY.
FT   DISULFID   439    448       BY SIMILARITY.
FT   DISULFID   455    466       BY SIMILARITY.
FT   DISULFID   460    475       BY SIMILARITY.
FT   DISULFID   477    486       BY SIMILARITY.
FT   DISULFID   493    504       BY SIMILARITY.
FT   DISULFID   498    513       BY SIMILARITY.
FT   DISULFID   515    524       BY SIMILARITY.
FT   DISULFID   531    542       BY SIMILARITY.
FT   DISULFID   536    551       BY SIMILARITY.
FT   DISULFID   553    562       BY SIMILARITY.
FT   DISULFID   569    579       BY SIMILARITY.
FT   DISULFID   574    588       BY SIMILARITY.
FT   DISULFID   590    599       BY SIMILARITY.
FT   DISULFID   606    617       BY SIMILARITY.
FT   DISULFID   611    626       BY SIMILARITY.
FT   DISULFID   628    637       BY SIMILARITY.
FT   DISULFID   644    654       BY SIMILARITY.
FT   DISULFID   649    663       BY SIMILARITY.
FT   DISULFID   665    674       BY SIMILARITY.
FT   DISULFID   681    692       BY SIMILARITY.
FT   DISULFID   686    701       BY SIMILARITY.
FT   DISULFID   703    712       BY SIMILARITY.
FT   DISULFID   719    729       BY SIMILARITY.
FT   DISULFID   724    738       BY SIMILARITY.
FT   DISULFID   740    749       BY SIMILARITY.
FT   DISULFID   756    767       BY SIMILARITY.
FT   DISULFID   761    776       BY SIMILARITY.
FT   DISULFID   778    787       BY SIMILARITY.
FT   DISULFID   794    805       BY SIMILARITY.
FT   DISULFID   799    814       BY SIMILARITY.
FT   DISULFID   816    825       BY SIMILARITY.
FT   DISULFID   832    843       BY SIMILARITY.
FT   DISULFID   837    854       BY SIMILARITY.
FT   DISULFID   856    865       BY SIMILARITY.
FT   DISULFID   872    883       BY SIMILARITY.
FT   DISULFID   877    892       BY SIMILARITY.
FT   DISULFID   894    903       BY SIMILARITY.
FT   DISULFID   910    921       BY SIMILARITY.
FT   DISULFID   915    930       BY SIMILARITY.
FT   DISULFID   932    941       BY SIMILARITY.
FT   DISULFID   986    997       BY SIMILARITY.
FT   DISULFID   991   1006       BY SIMILARITY.
FT   DISULFID  1008   1017       BY SIMILARITY.
FT   DISULFID  1024   1035       BY SIMILARITY.
FT   DISULFID  1029   1044       BY SIMILARITY.
FT   DISULFID  1046   1055       BY SIMILARITY.
FT   DISULFID  1062   1073       BY SIMILARITY.
FT   DISULFID  1067   1082       BY SIMILARITY.
FT   DISULFID  1084   1093       BY SIMILARITY.
FT   DISULFID  1100   1121       BY SIMILARITY.
FT   DISULFID  1115   1130       BY SIMILARITY.
FT   DISULFID  1132   1141       BY SIMILARITY.
FT   DISULFID  1148   1159       BY SIMILARITY.
FT   DISULFID  1153   1168       BY SIMILARITY.
FT   DISULFID  1170   1179       BY SIMILARITY.
FT   DISULFID  1186   1197       BY SIMILARITY.
FT   DISULFID  1191   1206       BY SIMILARITY.
FT   DISULFID  1208   1217       BY SIMILARITY.
FT   DISULFID  1224   1243       BY SIMILARITY.
FT   DISULFID  1237   1252       BY SIMILARITY.
FT   DISULFID  1254   1263       BY SIMILARITY.
FT   DISULFID  1270   1283       BY SIMILARITY.
FT   DISULFID  1275   1292       BY SIMILARITY.
FT   DISULFID  1294   1303       BY SIMILARITY.
FT   DISULFID  1310   1321       BY SIMILARITY.
FT   DISULFID  1315   1333       BY SIMILARITY.
FT   DISULFID  1335   1344       BY SIMILARITY.
FT   DISULFID  1351   1362       BY SIMILARITY.
FT   DISULFID  1356   1371       BY SIMILARITY.
FT   DISULFID  1373   1382       BY SIMILARITY.
FT   DISULFID  1390   1401       BY SIMILARITY.
FT   DISULFID  1395   1412       BY SIMILARITY.
FT   DISULFID  1414   1423       BY SIMILARITY.
FT   CARBOHYD   462    462       POTENTIAL.
FT   CARBOHYD   887    887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          10.0%;  Score 542;  DB 1;  Length 2524;
  Best Local Similarity  37.3%;  Pred. No. 2.61e-93;
  Matches   95;  Conservative   53;  Mismatches  76;  Indels  31;  Gaps  15;

Db    415 CSLGAN-PCEHGGRCTNTLGS-FQCNCPQGYAGPRCEIDVNECLSNPCQNDSTC-LDQIG 471
          | |  | || : : |  | :  : |||  |: | :|| ::: | ::|| |::|| : | |
Qy    180 CDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAG 239

Db    472 EFQCICMPGYEGLYCETNIDECASNPCLHNGKCID---------K------INEFRCDCP 516
          | | |  |:|| ||| |||:| :: | : |||:|         |      || :|||||
Qy    240 RFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCP 299

Db    517 TGFSGNLCQHDFDECTST--PCKNGAKCLDGPNSYTCQCTEGFTGRHCEQDINECIPDPC 574
            : |: |:  :: ||    || | :||:    ||:| |: |||| :|| :|::|    |
Qy    300 MEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVEC 359

Db    575 HYG-TCKDGIATFTCLCRPGYTGRLCD-NDI-N-ECL-SKPCL-N--G-GQCTDRENG-- 623
          : | :| ||| :: |||||||:|: |:   : : |   : :|  :  | |:|   :|:
Qy    360 QNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSD 419

Db    624 YICTCPKGTTGVNCE 638
          : | |  | :| :|:
Qy    420 FTCKCHEGFSGPSCD 434


RESULT   7
ID   DLL1_MOUSE     STANDARD;      PRT;   722 AA.
AC   Q61483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C X C57BL/6; TISSUE=EMBRYO;
```

```
RX   MEDLINE; 95401858.
RA   BETTENHAUSEN B., DE ANGELIS M.H., SIMON D., GUENET J.-L., GOSSLER A.;
RT   "Transient and restricted expression during mouse embryogenesis of
RT   Dll1, a murine gene closely related to Drosophila Delta.";
RL   DEVELOPMENT 121:2407-2418(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN THE PARAXIAL
CC       MESODERM AND NERVOUS SYSTEM. EXPRESSED AT HIGH LEVELS IN ADULT
CC       HEART AND AT LOWER LEVELS, IN ADULT LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED UNTIL DAY 15 IN THE EMBRYO.
CC       EXPRESSION THEN DECREASES AND INCREASES AGAIN IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80903; G806570; -.
DR   MGD; MGI:104659; DLL1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   722       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   722       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225   253       EGF-LIKE 1.
FT   DOMAIN     256   284       EGF-LIKE 2.
FT   DOMAIN     291   324       EGF-LIKE 3.
FT   DOMAIN     331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369   401       EGF-LIKE 5.
FT   DOMAIN     408   439       EGF-LIKE 6.
FT   DOMAIN     446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484   515       EGF-LIKE 8.
FT   DISULFID   225   236       BY SIMILARITY.
FT   DISULFID   229   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   262   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   303       BY SIMILARITY.
FT   DISULFID   297   313       BY SIMILARITY.
FT   DISULFID   315   324       BY SIMILARITY.
FT   DISULFID   331   342       BY SIMILARITY.
FT   DISULFID   336   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   380       BY SIMILARITY.
FT   DISULFID   374   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   CARBOHYD   476   476       POTENTIAL.
SQ   SEQUENCE   722 AA;  78448 MW;  5A647702 CRC32;
```

```
  Query Match          9.7%;  Score 527;  DB 1;  Length 722;
  Best Local Similarity  35.6%;  Pred. No. 8.40e-90;
  Matches   89;  Conservative   54;  Mismatches  80;  Indels  27;  Gaps  20;

Db    294 HKPCRNGATCTNTGQGSYTCSCRPGYTGANCELEVDECAPSPCKNGASCTDLEDS-FSCT 352
          : ||:| | |  |:   |||:| ||: | :|| ::| |  ||| | |:|   : : |:|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    353 CPPGFYGKVCE--LSA-M-T-CADGP-CFNGGR-CSD---N--P-D-GGYTCHCPLGFSG 398
          |  || |  ||  :   : : | :|  | :  | ||:   |  : :  :| | ||: : |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    399 FNCEKKMDLCGS--SPCSNGAKCVDLGNSYLCRCQAGFSGRYCEDNVDDCASSPCANGGT 456
           :|| |:: |    :|| | :||: :  || | | :||:|  || |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    457 CRDSVNDFSCTCPPGYTGKNCS-APV-SR-CEHA-PCH-N--G-ATC--HQRGQRYMCEC 506
          | |:: : | | |||:|  |  :|:    : : :|: :  | : |   | :  : | |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKC 424

Db    507 AQGYGGPNCQ 516
           :|::||:|:
Qy    425 HEGFSGPSCD 434
```

```
RESULT    8
ID   DLL1_HUMAN     STANDARD;      PRT;   723 AA.
AC   O00548;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MANN R.S., GRAY G.E., HENRIQUE D., ISH-HOROWICZ D.,
RA   ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF003522; G2197069; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   723       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   723       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     226   254       EGF-LIKE 1.
FT   DOMAIN     257   285       EGF-LIKE 2.
FT   DOMAIN     292   325       EGF-LIKE 3.
```

```
FT   DOMAIN      332    363       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN      370    402       EGF-LIKE 5.
FT   DOMAIN      409    440       EGF-LIKE 6.
FT   DOMAIN      447    478       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      485    516       EGF-LIKE 8.
FT   DISULFID    226    237       BY SIMILARITY.
FT   DISULFID    230    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    257    268       BY SIMILARITY.
FT   DISULFID    263    274       BY SIMILARITY.
FT   DISULFID    276    285       BY SIMILARITY.
FT   DISULFID    292    304       BY SIMILARITY.
FT   DISULFID    298    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    391       BY SIMILARITY.
FT   DISULFID    393    402       BY SIMILARITY.
FT   DISULFID    409    420       BY SIMILARITY.
FT   DISULFID    414    429       BY SIMILARITY.
FT   DISULFID    431    440       BY SIMILARITY.
FT   DISULFID    447    467       BY SIMILARITY.
FT   DISULFID    469    478       BY SIMILARITY.
FT   DISULFID    485    496       BY SIMILARITY.
FT   DISULFID    490    505       BY SIMILARITY.
FT   DISULFID    507    516       BY SIMILARITY.
FT   CARBOHYD    477    477       POTENTIAL.
SQ   SEQUENCE   723 AA;  77956 MW;  A1D48BDB CRC32;

  Query Match           9.7%;  Score 529;  DB 1;  Length 723;
  Best Local Similarity  35.2%;  Pred. No. 2.87e-90;
  Matches   88;  Conservative   56;  Mismatches  79;  Indels  27;  Gaps  20;

Db    295 HKPCKNGATCTNTGQGSYTCSCRPGYTGATCELGIDECDPSPCKNGGSCTDLENS-YSCT 353
          : |||| | |  |:   |||:| ||: |  ||  || |  ||| | ::|   : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    354 CPPGFYGKICE--LSA-M-T-CADGP-CFNGGR-CSD---S--P-D-GGYSCRCPVGYSG 399
          |  || |  ||  :   : : | :|  | :  | ||:   :  : :  :| | ||: | |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    400 FNCEKKIDYCSS--SPCSNGAKCVDLGDAYLCRCQAGFSGRHCDDNVDDCASSPCANGGT 457
          :|| |::||:   :|| | :||: :  :| | | :||:| :|: |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    458 CRDGVNDFSCTCPPGYTGRNCS-APV-SR-CEHA-PCH-N--G-ATC--HERGHGYVCEC 507
          | ||:  : | | |||:|: |  :|:    : : :|: :  | : |   : :  : | |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKC 424

Db    508 ARGYGGPNCQ 517
            |::||:|:
Qy    425 HEGFSGPSCD 434


RESULT    9
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL        1     90
FT   CHAIN        91   2139       CRUMBS PROTEIN.
FT   DOMAIN       91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2085   2111       POTENTIAL.
FT   DOMAIN     2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      267    303       EGF-LIKE 1.
FT   DOMAIN      306    343       EGF-LIKE 2.
FT   DOMAIN      348    386       EGF-LIKE 3.
FT   DOMAIN      388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      427    463       EGF-LIKE 5.
FT   DOMAIN      464    500       EGF-LIKE 6.
FT   DOMAIN      501    532       EGF-LIKE 7.
FT   DOMAIN      545    581       EGF-LIKE 8.
FT   DOMAIN      582    611       EGF-LIKE 9.
FT   DOMAIN      609    646       EGF-LIKE 10.
FT   DOMAIN      648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      840    902       EGF-LIKE 16.
FT   DOMAIN      904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      980   1021       EGF-LIKE 19.
FT   DOMAIN     1207   1243       EGF-LIKE 20.
FT   DOMAIN     1481   1517       EGF-LIKE 21.
FT   DOMAIN     1759   1795       EGF-LIKE 22.
FT   DOMAIN     1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1874   1915       EGF-LIKE 25.
FT   DOMAIN     1915   1951       EGF-LIKE 26.
```

```
FT   DOMAIN     1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     2030   2070       EGF-LIKE 29.
FT   DISULFID    271    282       BY SIMILARITY.
FT   DISULFID    276    291       BY SIMILARITY.
FT   DISULFID    293    302       BY SIMILARITY.
FT   DISULFID    310    321       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    333    342       BY SIMILARITY.
FT   DISULFID    352    363       BY SIMILARITY.
FT   DISULFID    357    374       BY SIMILARITY.
FT   DISULFID    376    385       BY SIMILARITY.
FT   DISULFID    392    403       BY SIMILARITY.
FT   DISULFID    397    412       BY SIMILARITY.
FT   DISULFID    414    424       BY SIMILARITY.
FT   DISULFID    431    442       BY SIMILARITY.
FT   DISULFID    436    451       BY SIMILARITY.
FT   DISULFID    453    462       BY SIMILARITY.
FT   DISULFID    468    479       BY SIMILARITY.
FT   DISULFID    473    488       BY SIMILARITY.
FT   DISULFID    490    499       BY SIMILARITY.
FT   DISULFID    505    515       BY SIMILARITY.
FT   DISULFID    509    520       BY SIMILARITY.
FT   DISULFID    522    531       BY SIMILARITY.
FT   DISULFID    549    562       BY SIMILARITY.
FT   DISULFID    556    569       BY SIMILARITY.
FT   DISULFID    571    580       BY SIMILARITY.
FT   DISULFID    586    597       BY SIMILARITY.
FT   DISULFID    591    602       BY SIMILARITY.
FT   DISULFID    604    610       BY SIMILARITY.
FT   DISULFID    613    624       BY SIMILARITY.
FT   DISULFID    618    634       BY SIMILARITY.
FT   DISULFID    636    645       BY SIMILARITY.
FT   DISULFID    652    664       BY SIMILARITY.
FT   DISULFID    659    673       BY SIMILARITY.
FT   DISULFID    675    684       BY SIMILARITY.
FT   DISULFID    691    702       BY SIMILARITY.
FT   DISULFID    696    711       BY SIMILARITY.
FT   DISULFID    713    722       BY SIMILARITY.
FT   DISULFID    729    740       BY SIMILARITY.
FT   DISULFID    734    749       BY SIMILARITY.
FT   DISULFID    751    760       BY SIMILARITY.
FT   DISULFID    767    778       BY SIMILARITY.
FT   DISULFID    772    787       BY SIMILARITY.
FT   DISULFID    789    799       BY SIMILARITY.
FT   DISULFID    806    817       BY SIMILARITY.
FT   DISULFID    811    826       BY SIMILARITY.
FT   DISULFID    828    837       BY SIMILARITY.
FT   DISULFID    844    855       BY SIMILARITY.
FT   DISULFID    849    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID    908    919       BY SIMILARITY.
FT   DISULFID    913    928       BY SIMILARITY.
FT   DISULFID    930    939       BY SIMILARITY.
FT   DISULFID    946    957       BY SIMILARITY.
FT   DISULFID    952    966       BY SIMILARITY.
FT   DISULFID    968    977       BY SIMILARITY.
FT   DISULFID    984    995       BY SIMILARITY.
FT   DISULFID    989   1009       BY SIMILARITY.
FT   DISULFID   1011   1020       BY SIMILARITY.
FT   DISULFID   1211   1222       BY SIMILARITY.
FT   DISULFID   1216   1231       BY SIMILARITY.
FT   DISULFID   1233   1242       BY SIMILARITY.
FT   DISULFID   1485   1496       BY SIMILARITY.
FT   DISULFID   1490   1505       BY SIMILARITY.
FT   DISULFID   1507   1516       BY SIMILARITY.
FT   DISULFID   1763   1774       BY SIMILARITY.
FT   DISULFID   1768   1783       BY SIMILARITY.
FT   DISULFID   1785   1794       BY SIMILARITY.
FT   DISULFID   1801   1812       BY SIMILARITY.
FT   DISULFID   1806   1821       BY SIMILARITY.
FT   DISULFID   1823   1832       BY SIMILARITY.
FT   DISULFID   1839   1850       BY SIMILARITY.
FT   DISULFID   1844   1859       BY SIMILARITY.
FT   DISULFID   1861   1870       BY SIMILARITY.
FT   DISULFID   1878   1889       BY SIMILARITY.
FT   DISULFID   1883   1903       BY SIMILARITY.
FT   DISULFID   1905   1914       BY SIMILARITY.
FT   DISULFID   1919   1930       BY SIMILARITY.
FT   DISULFID   1924   1939       BY SIMILARITY.
FT   DISULFID   1941   1950       BY SIMILARITY.
FT   DISULFID   1957   1968       BY SIMILARITY.
FT   DISULFID   1962   1977       BY SIMILARITY.
FT   DISULFID   1979   1988       BY SIMILARITY.
FT   DISULFID   1995   2008       BY SIMILARITY.
FT   DISULFID   2002   2017       BY SIMILARITY.
FT   DISULFID   2019   2028       BY SIMILARITY.
FT   CARBOHYD     37     37       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    238    238       POTENTIAL.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    336    336       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CARBOHYD    550    550       POTENTIAL.
FT   CARBOHYD    565    565       POTENTIAL.
FT   CARBOHYD    736    736       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    860    860       POTENTIAL.
FT   CARBOHYD    884    884       POTENTIAL.
FT   CARBOHYD    976    976       POTENTIAL.
FT   CARBOHYD   1102   1102       POTENTIAL.
FT   CARBOHYD   1114   1114       POTENTIAL.
FT   CARBOHYD   1138   1138       POTENTIAL.
FT   CARBOHYD   1192   1192       POTENTIAL.
FT   CARBOHYD   1245   1245       POTENTIAL.
FT   CARBOHYD   1255   1255       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   CARBOHYD   1363   1363       POTENTIAL.
FT   CARBOHYD   1441   1441       POTENTIAL.
FT   CARBOHYD   1454   1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match           9.7%;  Score 529;  DB 1;  Length 2139;
  Best Local Similarity  30.5%;  Pred. No. 2.87e-90;
  Matches  100;  Conservative   80;  Mismatches 116;  Indels  32;  Gaps  25;

Db    747 FTCNC-IPGMRGRICDIDIDDCVGDPCLNGGQCIDQLG-GFRCDCSGTGYEGENCELNID 804
          |||: :| : : ||:    |: || | : |    :  : |:|: |: | :||  ||
Qy    166 FTCDSKVPTKLATKCDL----CLNSPCKNNAICETTSSRKYTCNCT-PGFYGVHCENQID 220

Db    805 ECLSNPCTNGAKC-LDRVKDYFCDCHNGYKGKNCEQDINECESNPCQYNGNCLERSNITL 863
          | ::|| | | | : :   : | |::|: |  || :|::| :: |:  |:|::   :
Qy    221 ACYGSPCLNNATCKVAQAGRFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDL--VR- 277

Db    864 YQMSRITDLPKVFSQPFSFENASGYECVCVPGIIGKNCEININECDS--NPCSKHGNCND 921
          :     :| | | |:| : | | | | |     ||:|| ::: |    ||| ::|:|
Qy    278 F--CS-EEL-KNF-QSFQI-N-S-YRCDCPMEYEGKHCEDKLEYCTKKLNPCENNGKCIP 329

Db    922 GIGTYTCECEPGFEGTHCEINIDECDRYNPCQRG-TCYDQIDDYDCDCDANYGGKNCSV- 979
            |:|:| | ||| | :|| |||:| :   || | :| | |  ||| | : |:|  | :
Qy    330 INGSYSCMCSPGFTGNNCETNIDDC-KNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIP 388

Db    980 -LL-KGCDQ-NPCLNGGAC-LP-YLINEVTHLYTCTCENGFQGDKCEKTTTLS-MVATSL 1033
           :: :  :  ::|  : ||     : :: :  :|| | :|| |  |::  ::: : : :
Qy    389 PMMDMEYQKTDACQQS-ACGQGECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAY 447

Db   1034 ISVTTEREEGYDINLQFRTTLPNGVLAF 1061
           :::     :|  | : :|||   |:| :
Qy    448 LALDPLASDG-TITMTLRTTSKIGILLY 474


RESULT  10
ID   DL_DROME       STANDARD;      PRT;   880 AA.
AC   P10041;
```

```
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS DELTA PROTEIN PRECURSOR.
GN   DL.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   VAESSIN H., BREMER K.A., KNUST E., CAMPOS-ORTEGA J.A.;
RT   "The neurogenic gene Delta of Drosophila melanogaster is expressed in
RT   neurogenic territories and encodes a putative transmembrane protein
RT   with EGF-like repeats.";
RL   EMBO J. 6:3431-3440(1987).
RN   [2]
RP   SEQUENCE OF 422-621 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D., VAESSIN H.,
RA   CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
RN   [3]
RP   PATTERN OF TRANSCRIPTION.
RX   MEDLINE; 91209246.
RA   HAENLIN M., KRAMATSCHEK B., CAMPOS-ORTEGA J.A.;
RT   "The pattern of transcription of the neurogenic gene Delta of
RT   Drosophila melanogaster.";
RL   DEVELOPMENT 110:905-914(1990).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL
CC       IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL
CC       CELL LINEAGES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN LEVEL, IT IS
CC       CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY COMPETE
CC       FOR BINDING WITH THE NOTCH PROTEIN.
CC   -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA SERRATE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06289; G7853; -.
DR   EMBL; X05140; G929563; -.
DR   PIR; S00670; S00670.
DR   PIR; A26637; A26637.
DR   FLYBASE; FBgn0000463; Dl.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 8.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; TRANSMEMBRANE;
KW   EGF-LIKE DOMAIN; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   880       DELTA PROTEIN.
FT   DOMAIN      19   653       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   654   677       POTENTIAL.
FT   DOMAIN     678   880       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     227   258       EGF-LIKE 1.
FT   DOMAIN     256   289       EGF-LIKE 2.
FT   DOMAIN     291   329       EGF-LIKE 3.
FT   DOMAIN     331   372       EGF-LIKE 4.
FT   DOMAIN     374   416       EGF-LIKE 5.
FT   DOMAIN     418   451       EGF-LIKE 6.
FT   DOMAIN     453   489       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     491   527       EGF-LIKE 8.
FT   DOMAIN     529   565       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID   231   240       BY SIMILARITY.
FT   DISULFID   235   246       BY SIMILARITY.
FT   DISULFID   248   257       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   266   277       BY SIMILARITY.
FT   DISULFID   279   288       BY SIMILARITY.
FT   DISULFID   295   307       BY SIMILARITY.
FT   DISULFID   301   317       BY SIMILARITY.
FT   DISULFID   319   328       BY SIMILARITY.
FT   DISULFID   335   348       BY SIMILARITY.
FT   DISULFID   342   360       BY SIMILARITY.
FT   DISULFID   362   371       BY SIMILARITY.
FT   DISULFID   378   388       BY SIMILARITY.
FT   DISULFID   383   404       BY SIMILARITY.
FT   DISULFID   406   415       BY SIMILARITY.
FT   DISULFID   422   433       BY SIMILARITY.
FT   DISULFID   427   439       BY SIMILARITY.
FT   DISULFID   441   450       BY SIMILARITY.
FT   DISULFID   457   468       BY SIMILARITY.
FT   DISULFID   462   477       BY SIMILARITY.
FT   DISULFID   479   488       BY SIMILARITY.
FT   DISULFID   495   506       BY SIMILARITY.
FT   DISULFID   500   515       BY SIMILARITY.
FT   DISULFID   517   526       BY SIMILARITY.
FT   DISULFID   533   544       BY SIMILARITY.
FT   DISULFID   538   553       BY SIMILARITY.
FT   DISULFID   555   564       BY SIMILARITY.
FT   CARBOHYD    98    98       POTENTIAL.
FT   CARBOHYD   137   137       POTENTIAL.
FT   CARBOHYD   167   167       POTENTIAL.
FT   CARBOHYD   649   649       POTENTIAL.
FT   CONFLICT   437   438       GK -> ET (IN REF. 2).
FT   CONFLICT   459   459       G -> A (IN REF. 2).
FT   CONFLICT   490   490       S -> T (IN REF. 2).
SQ   SEQUENCE   880 AA;  94643 MW;  E967E662 CRC32;

  Query Match           9.6%;  Score 520;  DB 1;  Length 880;
  Best Local Similarity  37.1%;  Pred. No. 3.61e-88;
  Matches   95;  Conservative   55;  Mismatches  77;  Indels  29;  Gaps  22;

Db    295 CTNHRPCKNGGTCFNTGEGLYTCKCAPGYSGDDCENEIYSCDADVNPCQNGGTCIDEPHT 354
          | |  |||| : |  |:   |||:|:||: |  |||:| :| :  :|| | :||   :::
Qy    183 CLN-SPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYG--SPCLNNATC-KVAQA 238

Db    355 KTGYKCHCRNGWSGKMCEEKVLTCSDKPCHQG-ICRN-VR--PG-LGS-KG-Q--GYQCE 405
            ::| | :|  |  || ::  | :  |  |  | : ||  :  | :  : |  :|:|:
Qy    239 GR-FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCD 297

Db    406 CPIGYSGPNCDLQLDNCSP--NPCINGGSCQP-SGK--CICPSGFSGTRCETNIDDCLGH 460
          ||: | | :|:  |: |:   ||| | | | | :|   |:|::||:|  ||||||||
Qy    298 CPMEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNV 357

Db    461 QCENGGTCIDMVNQYRCQCVPGFHGTHC--SSKVDL-CL-IRPC--AN-G-GTCL-NLNN 511
          :|:|||:|:| :  | | | ||: |  |  ::::|:     :|  :  | | |: : |:
Qy    358 ECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNS 417

Db    512 -DYQCTCRAGFTGKDC 526
           |: | |: ||:|  |
Qy    418 SDFTCKCHEGFSGPSC 433


RESULT  11
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
```

```
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19  2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19  1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724  1746       POTENTIAL.
FT   DOMAIN    1747  2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    58       EGF-LIKE 1.
FT   DOMAIN      59    99       EGF-LIKE 2.
FT   DOMAIN     102   139       EGF-LIKE 3.
FT   DOMAIN     140   176       EGF-LIKE 4.
FT   DOMAIN     178   216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218   255       EGF-LIKE 6.
FT   DOMAIN     257   293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295   333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335   371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372   410       EGF-LIKE 10.
FT   DOMAIN     412   450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452   488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490   526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528   564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566   601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603   639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641   676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678   714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716   751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753   789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791   827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829   867       EGF-LIKE 22.
FT   DOMAIN     869   905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907   943       EGF-LIKE 24.
FT   DOMAIN     945   981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983  1019       EGF-LIKE 26.
FT   DOMAIN    1021  1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059  1095       EGF-LIKE 28.
FT   DOMAIN    1097  1143       EGF-LIKE 29.
FT   DOMAIN    1145  1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183  1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221  1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267  1305       EGF-LIKE 33.
FT   DOMAIN    1307  1346       EGF-LIKE 34.
FT   DOMAIN    1348  1384       EGF-LIKE 35.
FT   DOMAIN    1387  1426       EGF-LIKE 36.
FT   DOMAIN    1449  1462       CYS-RICH.
FT   DOMAIN    1865  2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865  1910       ANK MOTIF 1.
FT   REPEAT    1912  1942       ANK MOTIF 2.
FT   REPEAT    1944  1975       ANK MOTIF 3.
FT   REPEAT    1978  2009       ANK MOTIF 4.
FT   REPEAT    2011  2042       ANK MOTIF 5.
FT   REPEAT    2044  2076       ANK MOTIF 6.
FT   DISULFID    24    37       BY SIMILARITY.
FT   DISULFID    31    46       BY SIMILARITY.
FT   DISULFID    48    57       BY SIMILARITY.
FT   DISULFID    63    74       BY SIMILARITY.
FT   DISULFID    68    87       BY SIMILARITY.
FT   DISULFID    89    98       BY SIMILARITY.
FT   DISULFID   106   117       BY SIMILARITY.
FT   DISULFID   111   127       BY SIMILARITY.
FT   DISULFID   129   138       BY SIMILARITY.
FT   DISULFID   144   155       BY SIMILARITY.
FT   DISULFID   149   164       BY SIMILARITY.
FT   DISULFID   166   175       BY SIMILARITY.
FT   DISULFID   182   195       BY SIMILARITY.
FT   DISULFID   189   204       BY SIMILARITY.
FT   DISULFID   206   215       BY SIMILARITY.
FT   DISULFID   222   233       BY SIMILARITY.
FT   DISULFID   227   243       BY SIMILARITY.
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   261   272       BY SIMILARITY.
FT   DISULFID   266   281       BY SIMILARITY.
FT   DISULFID   283   292       BY SIMILARITY.
FT   DISULFID   299   312       BY SIMILARITY.
FT   DISULFID   306   321       BY SIMILARITY.
FT   DISULFID   323   332       BY SIMILARITY.
FT   DISULFID   339   350       BY SIMILARITY.
FT   DISULFID   344   359       BY SIMILARITY.
FT   DISULFID   361   370       BY SIMILARITY.
FT   DISULFID   376   387       BY SIMILARITY.
FT   DISULFID   381   398       BY SIMILARITY.
FT   DISULFID   400   409       BY SIMILARITY.
FT   DISULFID   416   429       BY SIMILARITY.
FT   DISULFID   423   438       BY SIMILARITY.
FT   DISULFID   440   449       BY SIMILARITY.
FT   DISULFID   456   467       BY SIMILARITY.
FT   DISULFID   461   476       BY SIMILARITY.
FT   DISULFID   478   487       BY SIMILARITY.
FT   DISULFID   494   505       BY SIMILARITY.
FT   DISULFID   499   514       BY SIMILARITY.
FT   DISULFID   516   525       BY SIMILARITY.
FT   DISULFID   532   543       BY SIMILARITY.
FT   DISULFID   537   552       BY SIMILARITY.
FT   DISULFID   554   563       BY SIMILARITY.
FT   DISULFID   570   580       BY SIMILARITY.
FT   DISULFID   575   589       BY SIMILARITY.
FT   DISULFID   591   600       BY SIMILARITY.
FT   DISULFID   607   618       BY SIMILARITY.
FT   DISULFID   612   627       BY SIMILARITY.
FT   DISULFID   629   638       BY SIMILARITY.
FT   DISULFID   645   655       BY SIMILARITY.
FT   DISULFID   650   664       BY SIMILARITY.
FT   DISULFID   666   675       BY SIMILARITY.
FT   DISULFID   682   693       BY SIMILARITY.
FT   DISULFID   687   702       BY SIMILARITY.
FT   DISULFID   704   713       BY SIMILARITY.
```

```
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    866       BY SIMILARITY.
FT   DISULFID    873    884       BY SIMILARITY.
FT   DISULFID    878    893       BY SIMILARITY.
FT   DISULFID    895    904       BY SIMILARITY.
FT   DISULFID    911    922       BY SIMILARITY.
FT   DISULFID    916    931       BY SIMILARITY.
FT   DISULFID    933    942       BY SIMILARITY.
FT   DISULFID    987    998       BY SIMILARITY.
FT   DISULFID    992   1007       BY SIMILARITY.
FT   DISULFID   1009   1018       BY SIMILARITY.
FT   DISULFID   1025   1036       BY SIMILARITY.
FT   DISULFID   1030   1045       BY SIMILARITY.
FT   DISULFID   1047   1056       BY SIMILARITY.
FT   DISULFID   1063   1074       BY SIMILARITY.
FT   DISULFID   1068   1083       BY SIMILARITY.
FT   DISULFID   1085   1094       BY SIMILARITY.
FT   DISULFID   1101   1122       BY SIMILARITY.
FT   DISULFID   1116   1131       BY SIMILARITY.
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.
FT   DISULFID   1225   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1271   1284       BY SIMILARITY.
FT   DISULFID   1276   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1311   1322       BY SIMILARITY.
FT   DISULFID   1316   1334       BY SIMILARITY.
FT   DISULFID   1336   1345       BY SIMILARITY.
FT   DISULFID   1352   1363       BY SIMILARITY.
FT   DISULFID   1357   1372       BY SIMILARITY.
FT   DISULFID   1374   1383       BY SIMILARITY.
FT   DISULFID   1391   1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           9.6%;  Score 520;  DB 1;  Length 2531;
  Best Local Similarity  37.3%;  Pred. No. 3.61e-88;
  Matches   94;  Conservative   48;  Mismatches  83;  Indels  27;  Gaps  15;

Db    757 CESNPCVNGGTCKDMTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTC-IDDVAGYK 814
          | ::|| | : |   :|  | | |  || | :|:  |: | ::|||| :|| : : : ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    815 CNCPLPYTGATCEVVLAPCATSPCKNSG-----V--CKES-EDYESF---S--CVCPTGW 861
          | |   : |  ||  :  |  | | |:|     |  | |   :::||   |  | ||
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    862 QGQTCEIDINECVK--SPCRHGASCQNTNGSYRCLCQAGYTGRNCESDIDDCRPNPCHNG 919
          :|  ||  :: | | |  :|| : : |   |||| |:| :|:|| |||::|||:   |:||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    920 GSCTDGVNAAFCDCLPGFQGAFCE-EDI-N-ECA-TNPCQNGA-NCTDCV---DS--YTC 969
          ||| ||: :  | | ||: | :||   : : |   |::|| :|    :||   :|  :||
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQRTDACQQSACGQGECVASQNSSDFTC 422

Db    970 TCPTGFNGIHCE 981
```

```
          | ||:|  |:
Qy    423 KCHEGFSGPSCD 434


RESULT   12
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
```

```
FT   REPEAT     1865   1910       ANK MOTIF 1.
FT   REPEAT     1912   1942       ANK MOTIF 2.
FT   REPEAT     1944   1975       ANK MOTIF 3.
FT   REPEAT     1978   2009       ANK MOTIF 4.
FT   REPEAT     2011   2042       ANK MOTIF 5.
FT   REPEAT     2044   2075       ANK MOTIF 6.
FT   CARBOHYD    888    888       POTENTIAL.
FT   CARBOHYD    959    959       POTENTIAL.
FT   CARBOHYD   1179   1179       POTENTIAL.
FT   CARBOHYD   1241   1241       POTENTIAL.
FT   CARBOHYD   1489   1489       POTENTIAL.
FT   CARBOHYD   1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match           9.5%;  Score 518;  DB 1;  Length 2531;
  Best Local Similarity  37.3%;  Pred. No. 1.06e-87;
  Matches   94;  Conservative   48;  Mismatches  83;  Indels  27;  Gaps  15;

Db    757 CESNPCVNGGTCKDMTS-GYVCTCREGFSGPNCQTNINECASNPCLNQGTC-IDDVAGYK 814
          | ::|| | : |   :|  | | |  || | :|:  |: | ::|||| :|| : : : ::
Qy    183 CLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFN 242

Db    815 CNCPLPYTPGATCEVVLAPCATSPCKNSG-----V--CKES-EDYESF---S--CVCPTGW 861
          | |   : |  ||  :  |  | | |:|     | | |   :::||   |  | ||
Qy    243 CYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEY 302

Db    862 QGQTCEVDINECVK--SPCRHGASCQNTNGSYRCLCQAGYTGRNCESDIDDCRPNPCHNG 919
          :|  ||  :: | |  :|| : : |   |||| |:| :|:|| |||::|||:   |:||
Qy    303 EGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNG 362

Db    920 GSCTDGINTAFCDCLPGFQGAFCE-EDI-N-ECA-SNPCQNGA-NCTDCV---DS--YTC 969
          ||| |||  :  | | ||: | :||   : : |   :::|| :|    :||   :|  :||
Qy    363 GSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTC 422

Db    970 TCPVGFNGIHCE 981
          |  ||:|  |:
Qy    423 KCHEGFSGPSCD 434


RESULT  13
ID   DLL1_RAT       STANDARD;      PRT;   714 AA.
AC   P97677;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DISIBIO G., HEBSHI L., BOULTER J., WEINMASTER G.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78889; G1699046; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    714       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18    537       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   538    560       POTENTIAL.
FT   DOMAIN     561    714       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225    253       EGF-LIKE 1.
FT   DOMAIN     256    284       EGF-LIKE 2.
FT   DOMAIN     291    324       EGF-LIKE 3.
FT   DOMAIN     331    362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369    401       EGF-LIKE 5.
FT   DOMAIN     408    439       EGF-LIKE 6.
FT   DOMAIN     446    477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484    515       EGF-LIKE 8.
FT   DISULFID   225    236       BY SIMILARITY.
FT   DISULFID   229    242       BY SIMILARITY.
FT   DISULFID   244    253       BY SIMILARITY.
FT   DISULFID   256    267       BY SIMILARITY.
FT   DISULFID   262    273       BY SIMILARITY.
FT   DISULFID   275    284       BY SIMILARITY.
FT   DISULFID   291    303       BY SIMILARITY.
FT   DISULFID   297    313       BY SIMILARITY.
FT   DISULFID   315    324       BY SIMILARITY.
FT   DISULFID   331    342       BY SIMILARITY.
FT   DISULFID   336    351       BY SIMILARITY.
FT   DISULFID   353    362       BY SIMILARITY.
FT   DISULFID   369    380       BY SIMILARITY.
FT   DISULFID   374    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   430    439       BY SIMILARITY.
FT   DISULFID   446    466       BY SIMILARITY.
FT   DISULFID   468    477       BY SIMILARITY.
FT   DISULFID   484    495       BY SIMILARITY.
FT   DISULFID   489    504       BY SIMILARITY.
FT   DISULFID   506    515       BY SIMILARITY.
FT   CARBOHYD   476    476       POTENTIAL.
SQ   SEQUENCE   714 AA;  77378 MW;  604B76D1 CRC32;

  Query Match           9.3%;  Score 508;  DB 1;  Length 714;
  Best Local Similarity  34.4%;  Pred. No. 2.25e-85;
  Matches   86;  Conservative   57;  Mismatches  80;  Indels  27;  Gaps  20;

Db    294 HKPCRNGATCTNTGQGSYTCSCRPGYTGANCELEVDECAPSPCRNGGSCTDLEDS-YSCT 352
          : ||:| | |  |:  |||:| ||: | :|| ::| |  ||| | ::|  : : ::|
Qy    185 NSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGRFNCY 244

Db    353 CPPGFYGKVCE--LSA-M-T-CADGP-CFNGGR-CSD---N--P-D-GGYTCHCPAGFSG 398
          |  || |  ||  :  : : | :|  | :  | ||:   |  : :  :| | ||  : |
Qy    245 CNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPMEYEG 304

Db    399 FNCEKKIDLCSS--SPCSNGAKCVDLGNSYLCRCQTGFSGRYCEDNVDDCASSPCANGGT 456
            :|| |:: |:   :|| | :||: :  || | |  ||:|  || |:||| :  | |||:
Qy    305 KHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGS 364

Db    457 CRDSVNDFSCTCPPGYTGRNCS-APV-SR-CEHA-PCH-N--G-ATC--HQRGQRYMCEC 506
          | |::  : | | |||:|: |  :|:     : : :|: :  | : |   | :  : | |
Qy    365 CVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKC 424

Db    507 AQGYGGANCQ 516
            :|::|::|:
Qy    425 HEGFSGPSCD 434


RESULT  14
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
```

```
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643      EXTRACELLULAR.
FT   TRANSMEM  1644   1664      POTENTIAL.
FT   DOMAIN    1665   2318      CYTOPLASMIC.
FT   DOMAIN      39   1374      34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503      3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998      6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261      PEST.
FT   DOMAIN      39     78      EGF-LIKE 1.
FT   DOMAIN      79    119      EGF-LIKE 2.
FT   DOMAIN     120    157      EGF-LIKE 3.
FT   DOMAIN     159    196      EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235      EGF-LIKE 5.
FT   DOMAIN     237    273      EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313      EGF-LIKE 7.
FT   DOMAIN     315    351      EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390      EGF-LIKE 9.
FT   DOMAIN     392    430      EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468      EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506      EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544      EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581      EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619      EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656      EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694      EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731      EGF-LIKE 18.
FT   DOMAIN     735    771      EGF-LIKE 19.
FT   DOMAIN     772    809      EGF-LIKE 20.
FT   DOMAIN     811    848      EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886      EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923      EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961      EGF-LIKE 24.
FT   DOMAIN     963    999      EGF-LIKE 25.
FT   DOMAIN    1001   1035      EGF-LIKE 26.
FT   DOMAIN    1037   1083      EGF-LIKE 27.
FT   DOMAIN    1085   1121      EGF-LIKE 28.
FT   DOMAIN    1123   1159      EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204      EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245      EGF-LIKE 31.
FT   DOMAIN    1247   1288      EGF-LIKE 32.
FT   DOMAIN    1290   1326      EGF-LIKE 33.
FT   DOMAIN    1336   1374      EGF-LIKE 34.
FT   REPEAT    1388   1428      LIN/NOTCH 1.
FT   REPEAT    1429   1467      LIN/NOTCH 2.
FT   REPEAT    1468   1503      LIN/NOTCH 3.
FT   REPEAT    1784   1816      CDC10/SWI6 1.
FT   REPEAT    1817   1865      CDC10/SWI6 2.
FT   REPEAT    1866   1898      CDC10/SWI6 3.
FT   REPEAT    1899   1932      CDC10/SWI6 4.
FT   REPEAT    1933   1965      CDC10/SWI6 5.
FT   REPEAT    1966   1998      CDC10/SWI6 6.
FT   DISULFID    43     55      BY SIMILARITY.
FT   DISULFID    49     66      BY SIMILARITY.
FT   DISULFID    68     77      BY SIMILARITY.
FT   DISULFID    83     94      BY SIMILARITY.
FT   DISULFID    88    107      BY SIMILARITY.
FT   DISULFID   109    118      BY SIMILARITY.
FT   DISULFID   124    135      BY SIMILARITY.
FT   DISULFID   129    145      BY SIMILARITY.
FT   DISULFID   147    156      BY SIMILARITY.
FT   DISULFID   163    175      BY SIMILARITY.
FT   DISULFID   169    184      BY SIMILARITY.
FT   DISULFID   186    195      BY SIMILARITY.
FT   DISULFID   202    213      BY SIMILARITY.
FT   DISULFID   207    223      BY SIMILARITY.
FT   DISULFID   225    234      BY SIMILARITY.
FT   DISULFID   241    252      BY SIMILARITY.
FT   DISULFID   246    261      BY SIMILARITY.
FT   DISULFID   263    272      BY SIMILARITY.
FT   DISULFID   279    292      BY SIMILARITY.
FT   DISULFID   286    301      BY SIMILARITY.
FT   DISULFID   303    312      BY SIMILARITY.
FT   DISULFID   319    330      BY SIMILARITY.
FT   DISULFID   324    339      BY SIMILARITY.
FT   DISULFID   341    350      BY SIMILARITY.
FT   DISULFID   356    367      BY SIMILARITY.
FT   DISULFID   361    378      BY SIMILARITY.
FT   DISULFID   380    389      BY SIMILARITY.
FT   DISULFID   396    409      BY SIMILARITY.
FT   DISULFID   403    418      BY SIMILARITY.
FT   DISULFID   420    429      BY SIMILARITY.
FT   DISULFID   436    447      BY SIMILARITY.
FT   DISULFID   441    456      BY SIMILARITY.
FT   DISULFID   458    467      BY SIMILARITY.
FT   DISULFID   474    485      BY SIMILARITY.
FT   DISULFID   479    494      BY SIMILARITY.
FT   DISULFID   496    505      BY SIMILARITY.
FT   DISULFID   512    523      BY SIMILARITY.
FT   DISULFID   517    532      BY SIMILARITY.
FT   DISULFID   534    543      BY SIMILARITY.
FT   DISULFID   550    560      BY SIMILARITY.
FT   DISULFID   555    569      BY SIMILARITY.
FT   DISULFID   571    580      BY SIMILARITY.
FT   DISULFID   587    598      BY SIMILARITY.
FT   DISULFID   592    607      BY SIMILARITY.
FT   DISULFID   609    618      BY SIMILARITY.
FT   DISULFID   625    635      BY SIMILARITY.
FT   DISULFID   630    644      BY SIMILARITY.
FT   DISULFID   646    655      BY SIMILARITY.
FT   DISULFID   662    673      BY SIMILARITY.
FT   DISULFID   667    682      BY SIMILARITY.
FT   DISULFID   684    693      BY SIMILARITY.
```

```
FT   DISULFID    700    710       BY SIMILARITY.
FT   DISULFID    705    719       BY SIMILARITY.
FT   DISULFID    721    730       BY SIMILARITY.
FT   DISULFID    739    750       BY SIMILARITY.
FT   DISULFID    744    759       BY SIMILARITY.
FT   DISULFID    761    770       BY SIMILARITY.
FT   DISULFID    776    787       BY SIMILARITY.
FT   DISULFID    781    797       BY SIMILARITY.
FT   DISULFID    799    808       BY SIMILARITY.
FT   DISULFID    815    827       BY SIMILARITY.
FT   DISULFID    821    836       BY SIMILARITY.
FT   DISULFID    838    847       BY SIMILARITY.
FT   DISULFID    854    865       BY SIMILARITY.
FT   DISULFID    859    874       BY SIMILARITY.
FT   DISULFID    876    885       BY SIMILARITY.
FT   DISULFID    892    902       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    913    922       BY SIMILARITY.
FT   DISULFID    929    940       BY SIMILARITY.
FT   DISULFID    934    949       BY SIMILARITY.
FT   DISULFID    951    960       BY SIMILARITY.
FT   DISULFID    967    978       BY SIMILARITY.
FT   DISULFID    972    987       BY SIMILARITY.
FT   DISULFID    989    998       BY SIMILARITY.
FT   DISULFID   1005   1016       BY SIMILARITY.
FT   DISULFID   1010   1023       BY SIMILARITY.
FT   DISULFID   1025   1034       BY SIMILARITY.
FT   DISULFID   1041   1062       BY SIMILARITY.
FT   DISULFID   1056   1071       BY SIMILARITY.
FT   DISULFID   1073   1082       BY SIMILARITY.
FT   DISULFID   1089   1100       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           9.3%;  Score 505;  DB 1;  Length 2318;
  Best Local Similarity  34.7%;  Pred. No. 1.12e-84;
  Matches   90;  Conservative   50;  Mismatches 103;  Indels  16;  Gaps  13;

Db    965 DPCFSRPCLHGGICNPTHP-GFECTCREGFTGSQCQNPVDWCSQAPCQNGGRC-V-QTGA 1021
          | |:: || : :||: | :  : | |  || | :|:| :| |  :|| | : | | |:|
Qy    181 DLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAGR 240

Db   1022 Y-CICPPGWSGRLCDIQSLPCTEAAAQMG---VRLEQLCQEGGKCIDKGRSH-Y-CVCPE 1075
          : | |  |  |  |:     | ::  : |  | | ::| |  | ::  : : | | ||
Qy    241 FNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCPM 300

Db   1076 GRTGSHCEHEVDPCTAQ--PCQHGGTCRGYMGGYVCECPAGYAGDSCEDNIDECASQPCQ 1133
             | |||  :: ||    ||:: | |    |:| | |::|::|::|| |||:| :  ||
Qy    301 EYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVECQ 360

Db   1134 NGGSCIDLVARYLCSCPPGTLGVLCEINEDDCDLG-PSLDSGVQCLHN-GTCVD-LVGG- 1189
          |||||:| :  | | | ||  | |||     |:     |: |     | ||    ::
Qy    361 NGGSCVDGILSYDCLCRPGYAGQYCEIPPMM-DMEYQKTDACQQSACGQGECVASQNSSD 419

Db   1190 FRCNCPPGYTGLHCEADIN 1208
          | |:|  |::|  |: :::
Qy    420 FTCKCHEGFSGPSCDRQMS 438


RESULT   15
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   >2444      NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737   1757       POTENTIAL.
FT   DOMAIN    1758   >2444      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
```

```
FT   DOMAIN      412    450      EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452    488      EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490    526      EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528    564      EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566    601      EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603    639      EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641    676      EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678    714      EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716    751      EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789      EGF-LIKE 20.
FT   DOMAIN      791    827      EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    868      EGF-LIKE 22.
FT   DOMAIN      870    906      EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      908    944      EGF-LIKE 24.
FT   DOMAIN      946    982      EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984   1020      EGF-LIKE 26.
FT   DOMAIN     1022   1058      EGF-LIKE 27.
FT   DOMAIN     1060   1096      EGF-LIKE 28.
FT   DOMAIN     1098   1144      EGF-LIKE 29.
FT   DOMAIN     1146   1182      EGF-LIKE 30.
FT   DOMAIN     1184   1220      EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1222   1266      EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1268   1306      EGF-LIKE 33.
FT   DOMAIN     1308   1347      EGF-LIKE 34.
FT   DOMAIN     1349   1385      EGF-LIKE 35.
FT   DOMAIN     1388   1427      EGF-LIKE 36.
FT   DOMAIN     1446   1563      3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1481      LIN/NOTCH 1.
FT   REPEAT     1482   1523      LIN/NOTCH 2.
FT   REPEAT     1524   1563      LIN/NOTCH 3.
FT   DOMAIN     1876   2087      6 X ANK MOTIF REPEATS.
FT   REPEAT     1876   1921      ANK MOTIF 1.
FT   REPEAT     1923   1954      ANK MOTIF 2.
FT   REPEAT     1956   1987      ANK MOTIF 3.
FT   REPEAT     1990   2021      ANK MOTIF 4.
FT   REPEAT     2023   2054      ANK MOTIF 5.
FT   REPEAT     2056   2087      ANK MOTIF 6.
FT   DOMAIN     1576   1579      POLY-VAL.
FT   DOMAIN     1662   1665      POLY-ARG.
FT   DOMAIN     1729   1732      POLY-PRO.
FT   DOMAIN     1741   1744      POLY-ALA.
FT   DOMAIN     1902   1905      POLY-GLU.
FT   DOMAIN     2260   2263      POLY-GLY.
FT   DOMAIN     2404   2407      POLY-GLN.
FT   DOMAIN     2411   2418      POLY-PRO.
FT   DISULFID     24     37      BY SIMILARITY.
FT   DISULFID     31     46      BY SIMILARITY.
FT   DISULFID     48     57      BY SIMILARITY.
FT   DISULFID     63     74      BY SIMILARITY.
FT   DISULFID     68     87      BY SIMILARITY.
FT   DISULFID     89     98      BY SIMILARITY.
FT   DISULFID    106    117      BY SIMILARITY.
FT   DISULFID    111    127      BY SIMILARITY.
FT   DISULFID    129    138      BY SIMILARITY.
FT   DISULFID    144    155      BY SIMILARITY.
FT   DISULFID    149    164      BY SIMILARITY.
FT   DISULFID    166    175      BY SIMILARITY.
FT   DISULFID    182    195      BY SIMILARITY.
FT   DISULFID    189    204      BY SIMILARITY.
FT   DISULFID    206    215      BY SIMILARITY.
FT   DISULFID    222    233      BY SIMILARITY.
FT   DISULFID    227    243      BY SIMILARITY.
FT   DISULFID    245    254      BY SIMILARITY.
FT   DISULFID    261    272      BY SIMILARITY.
FT   DISULFID    266    281      BY SIMILARITY.
FT   DISULFID    283    292      BY SIMILARITY.
FT   DISULFID    299    312      BY SIMILARITY.
FT   DISULFID    306    321      BY SIMILARITY.
FT   DISULFID    323    332      BY SIMILARITY.
FT   DISULFID    339    350      BY SIMILARITY.
FT   DISULFID    344    359      BY SIMILARITY.
FT   DISULFID    361    370      BY SIMILARITY.
FT   DISULFID    376    387      BY SIMILARITY.
FT   DISULFID    381    398      BY SIMILARITY.
FT   DISULFID    400    409      BY SIMILARITY.
FT   DISULFID    416    429      BY SIMILARITY.
FT   DISULFID    423    438      BY SIMILARITY.
FT   DISULFID    440    449      BY SIMILARITY.
FT   DISULFID    456    467      BY SIMILARITY.
FT   DISULFID    461    476      BY SIMILARITY.
FT   DISULFID    478    487      BY SIMILARITY.
FT   DISULFID    494    505      BY SIMILARITY.
FT   DISULFID    499    514      BY SIMILARITY.
FT   DISULFID    516    525      BY SIMILARITY.
FT   DISULFID    532    543      BY SIMILARITY.
FT   DISULFID    537    552      BY SIMILARITY.
FT   DISULFID    554    563      BY SIMILARITY.
FT   DISULFID    570    580      BY SIMILARITY.
FT   DISULFID    575    589      BY SIMILARITY.
FT   DISULFID    591    600      BY SIMILARITY.
FT   DISULFID    607    618      BY SIMILARITY.
FT   DISULFID    612    627      BY SIMILARITY.
FT   DISULFID    629    638      BY SIMILARITY.
FT   DISULFID    645    655      BY SIMILARITY.
FT   DISULFID    650    664      BY SIMILARITY.
FT   DISULFID    666    675      BY SIMILARITY.
FT   DISULFID    682    693      BY SIMILARITY.
FT   DISULFID    687    702      BY SIMILARITY.
FT   DISULFID    704    713      BY SIMILARITY.
FT   DISULFID    720    730      BY SIMILARITY.
FT   DISULFID    725    739      BY SIMILARITY.
FT   DISULFID    741    750      BY SIMILARITY.
FT   DISULFID    757    768      BY SIMILARITY.
FT   DISULFID    762    777      BY SIMILARITY.
FT   DISULFID    779    788      BY SIMILARITY.
FT   DISULFID    795    806      BY SIMILARITY.
FT   DISULFID    800    815      BY SIMILARITY.
FT   DISULFID    817    826      BY SIMILARITY.
FT   DISULFID    833    844      BY SIMILARITY.
FT   DISULFID    838    855      BY SIMILARITY.
FT   DISULFID    857    867      BY SIMILARITY.
FT   DISULFID    874    885      BY SIMILARITY.
FT   DISULFID    879    894      BY SIMILARITY.
FT   DISULFID    896    905      BY SIMILARITY.
FT   DISULFID    912    923      BY SIMILARITY.
FT   DISULFID    917    932      BY SIMILARITY.
FT   DISULFID    934    943      BY SIMILARITY.
FT   DISULFID    988    999      BY SIMILARITY.
FT   DISULFID    993   1008      BY SIMILARITY.
FT   DISULFID   1010   1019      BY SIMILARITY.
FT   DISULFID   1026   1037      BY SIMILARITY.
FT   DISULFID   1031   1046      BY SIMILARITY.
FT   DISULFID   1048   1057      BY SIMILARITY.
FT   DISULFID   1064   1075      BY SIMILARITY.
FT   DISULFID   1069   1084      BY SIMILARITY.
FT   DISULFID   1086   1095      BY SIMILARITY.
FT   DISULFID   1102   1123      BY SIMILARITY.
FT   DISULFID   1117   1132      BY SIMILARITY.
FT   DISULFID   1134   1143      BY SIMILARITY.
FT   DISULFID   1150   1161      BY SIMILARITY.
FT   DISULFID   1155   1170      BY SIMILARITY.
FT   DISULFID   1172   1181      BY SIMILARITY.
FT   DISULFID   1188   1199      BY SIMILARITY.
FT   DISULFID   1193   1208      BY SIMILARITY.
...
Note: remainder of annotations omitted.

 Query Match           9.3%;  Score 506;  DB 1;  Length 2444;
 Best Local Similarity  38.2%;  Pred. No. 6.56e-85;
 Matches   79;  Conservative   43;  Mismatches  64;  Indels  21;  Gaps  11;

Db    416 CSLGAN-PCEHAGKCINTLGS-FECQCLQGYTGPRCEIDVNECVSNPCQNDATC-LDQIG 472
          | |  | || : : |  | :  : | |  |: | :|| ::: | ::|| |:||| : | |
Qy    180 CDLCLNSPCKNNAICETTSSRKYTCNCTPGFYGVHCENQIDACYGSPCLNNATCKVAQAG 239

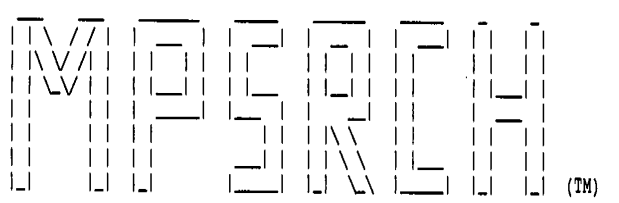
Db    473 EFQCMCMPGYEGVHCEVNTDECASSPCLHNGRC--L-----DKI-N-E-FQ-----CECP 517
```

```
          | | |  |:||  || | |:| :| | : |:|  |     : : | : ||     |:||
Qy    240 RFNCYCNKGFEGDYCEKNIDDCVNSKCENGGKCVDLVRFCSEELKNFQSFQINSYRCDCP 299

Db    518 TGFTGHLCQYDVDECAST--PCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPC 575
          : |  |:  :: |:    || | :||:    :|:|:|: |:|| :|| :||:|    |
Qy    300 MEYEGKHCEDKLEYCTKKLNPCENNGKCIPINGSYSCMCSPGFTGNNCETNIDDCKNVEC 359

Db    576 HYG-SCKDGVATFTCLCRPGYTGHHCE 601
          : | || ||: :: |||||||:|: ||
Qy    360 QNGGSCVDGILSYDCLCRPGYAGQYCE 386


Search completed: Fri May 28 09:12:50 1999
Job time : 59 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:18:21 1999; MasPar time 8.31 Seconds
394.061 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-10
Description: (1-154) from US09191647.pep
Perfect Score: 1160
Sequence: 1 DPLPVHHRCECMLGYTGDNC.....EDNGILLYNGDNDHIAVELY 154

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tatistics: Mean 29.725; Variance 145.891; scale 0.204

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1026	88.4	1534	30	W46966	Amino acid sequence o	1.17e-76
2	462	39.8	1480	5	R25079	Drosophila SLIT prote	1.72e-28
3	335	28.9	520	25	W18348	Proliferation and dif	5.96e-18
4	335	28.9	702	25	W18349	Proliferation and dif	5.96e-18
5	335	28.9	723	25	W18353	Proliferation and dif	5.96e-18
6	325	28.0	727	21	W11719	C-Delta-1 polypeptide	3.93e-17
7	325	28.0	740	21	W00876	C-Delta-1 polypeptide	3.93e-17
8	323	27.8	660	21	W11725	H-Delta-1 polypeptide	5.73e-17
9	321	27.7	722	21	W11720	M-Delta-1 polypeptide	8.35e-17
10	292	25.2	1036	25	W18351	Proliferation and dif	1.92e-14
11	292	25.2	1187	25	W18352	Proliferation and dif	1.92e-14
12	292	25.2	1218	25	W18354	Proliferation and dif	1.92e-14
13	292	25.2	1218	19	W05833	Human Serrate-1 (HJ1)	1.92e-14
14	292	25.2	1218	29	W44301	Human serrate 1.	1.92e-14
15	290	25.0	1193	19	W05835	Chick Serrate.	2.79e-14
16	287	24.7	1208	28	W40827	Human Jagged protein.	4.88e-14
17	280	24.1	2321	36	W49698	Human Notch3 protein.	1.80e-13
18	278	24.0	1872	36	W68510	Partial human Notch-3	2.61e-13
19	277	23.9	1055	29	W44298	Human serrate 2 prote	3.14e-13
20	277	23.9	1212	29	W44299	Human serrate 2.	3.14e-13
21	277	23.9	1257	19	W05834	Human Serrate-2 (HJ2)	3.14e-13
22	274	23.6	612	28	W39256	Human partial mature	5.48e-13
23	274	23.6	737	28	W39257	Human membrane protei	5.48e-13
24	271	23.4	2707	24	W27161	Mouse receptor ME2.	9.57e-13
25	265	22.8	685	37	W80813	Nucleotide sequence o	2.91e-12
26	265	22.8	1404	7	R38304	Sequence of a serrate	2.91e-12
27	253	21.8	157	21	W11730	H-Delta-1 polypeptide	2.66e-11
28	242	20.9	833	6	R28960	Delta D11.	2.01e-10
29	237	20.4	385	10	R56167	Neuroendocrine tumor	5.02e-10
30	228	19.7	1572	24	W27160	Mouse receptor ME2 re	2.60e-09
31	227	19.6	1257	9	R46627	Neurocan core protein	3.11e-09
32	222	19.1	383	10	R56166	Neuroendocrine tumor	7.73e-09
33	219	18.9	2409	3	R12609	Versican.	1.33e-08
34	218	18.8	473	17	R86869	Adhesive protein.	1.60e-08
35	210	18.1	77	6	R28962	ELR-11 and -12.	6.77e-08
36	205	17.7	4544	11	R60517	Human alpha-2-MR.	1.67e-07
37	205	17.7	4544	9	R47861	Alpha 2-Macroglobulin	1.67e-07
38	192	16.6	810	27	W37500	Human nel-related pro	1.71e-06
39	184	15.9	2189	1	R05222	Antigen GX5401FL enco	7.06e-06
40	181	15.6	179	37	W75100	Human secreted protei	1.20e-05
41	174	15.0	816	27	W37501	Human nel-related pro	4.12e-05
42	173	14.9	228	30	W46967	Amino acid sequence o	4.91e-05
43	172	14.8	240	33	W64219	Human secreted protei	5.85e-05
44	171	14.7	1833	14	R79478	Mouse LTBP-2.	6.97e-05
45	165	14.2	466	8	R52562	Factor VIII.	1.99e-04

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 88.4%; Score 1026; DB 30; Length 1534;
Best Local Similarity 91.3%; Pred. No. 1.17e-76;
Matches 137; Conservative 8; Mismatches 2; Indels 3; Gaps 1;

```
Db   1068 rcecmpgyagdncsenqddcrdhrcqngaqcmdevnsysclcaegysgqlceipphlpap 1127
          ||||| ||:||||||||||:||:|||||||:||||||:||| ||||||||||||   ||
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPP---AP 64

Db   1128 kspcegtecqngancvdqgnrpvcqclpgfggpecekllsvnfvdrdtylqftdlqnwpr 1187
          :|:|||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
```

```
Qy     65 RSSCEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPR 124

Db   1188 anitlqvstaedngillyngdndhiavely 1217
          ||||||||||||||||||||||||||||||
Qy    125 ANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match          39.8%;  Score 462;  DB 5;  Length 1480;
  Best Local Similarity  42.7%;  Pred. No. 1.72e-28;
  Matches   67;  Conservative   38;  Mismatches  44;  Indels   8;  Gaps   4;

Db   1047 hyscdcqagfhgtnctdniddcqnhmcqnggtcvdgindyqcrcpddytgkyceghnmis 1106
          |  |:|  |: | ||::| ||| :|:|||||: ||| :| | | | : |:|  ||    |:
Qy      6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db   1107 mmypqtspcqnheckhgvcfqpnaqgsdylcrchpgytgkwceyltsisfvhnnsfvele 1166
             |::| |:  || :|      |||  :|:| ||: |  || | |::||  ::::::
Qy     62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db   1167 plrtrpeanvtivfssaeqngilmydgqdahlavelf 1203
           |: :| ||:|:  |:||:||||:|:|:: |:||||:
Qy    118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT    3
ID   W18348 standard; protein; 520 AA.
AC   W18348;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 3; Page 59-61; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   520 AA;

  Query Match          28.9%;  Score 335;  DB 25;  Length 520;
  Best Local Similarity  44.8%;  Pred. No. 5.96e-18;
  Matches   43;  Conservative   20;  Mismatches  30;  Indels   3;  Gaps   2;

Db    408 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-apvsr--c 464
          | |  |::| :| :| |||    | ||: | | || ::| |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    465 ehapchngatcherghryvcecargyggpncqfllp 500
          | : |:||| | ::| | ||:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    4
ID   W18349 standard; protein; 702 AA.
AC   W18349;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
```

```
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 4; Page 61-64; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   702 AA;

  Query Match           28.9%;  Score 335;  DB 25;  Length 702;
  Best Local Similarity  44.8%;  Pred. No. 5.96e-18;
  Matches   43;  Conservative   20;  Mismatches  30;  Indels   3;  Gaps   2;

Db    408 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-apvsr--c 464
          | |  |::| :| :| |||    | ||: | | || ::| |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    465 ehapchngatcherghryvcecargyggpncqfllp 500
          | : |:||| | ::| | ||:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT   5
ID   W18353 standard; protein; 723 AA.
AC   W18353;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Peptide        1..21
FT                   /label= Signal
FT   Protein        22..723
FT                   /label= Differentiation_suppression_protein
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
DR   N-PSDB; T70174.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 15; Page 77-82; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   723 AA;

  Query Match           28.9%;  Score 335;  DB 25;  Length 723;
  Best Local Similarity  44.8%;  Pred. No. 5.96e-18;
  Matches   43;  Conservative   20;  Mismatches  30;  Indels   3;  Gaps   2;

Db    429 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-apvsr--c 485
          | |  |::| :| :| |||    | ||: | | || ::| |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    486 ehapchngatcherghryvcecargyggpncqfllp 521
          | : |:||| | ::| | ||:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT   6
ID   W11719 standard; Protein; 727 AA.
AC   W11719;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide.
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key            Location/Qualifiers
FT   domain         184..228
FT                   /label= DSL
FT   domain         229..261
FT                   /label= EGF1
FT   domain         262..292
FT                   /label= EGF2
FT   domain         293..332
FT                   /label= EGF3
FT   domain         333..370
FT                   /label= EGF4
FT   domain         371..409
FT                   /label= EGF5
FT   domain         410..447
FT                   /label= EGF6
FT   domain         448..485
FT                   /label= EGF7
FT   domain         486..523
FT                   /label= EGF8
FT   domain         524..534
FT                   /label= EGF9
FT   domain         555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58897.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W11719) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
CC   C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
CC   amino acid sequence was deduced from a cDNA clone (T58897) obtd.
CC   from chick stage 4-6 embryos.  An alternatively spliced variant
CC   (W00876) was also isolated, and mouse (W11720) and human (W11721-
CC   38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC   can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
CC   cancer, melanoma or seminoma, and nervous system disorders or to
CC   promote tissue regeneration and repair.
SQ   Sequence   727 AA;

  Query Match           28.0%;  Score 325;  DB 21;  Length 727;
  Best Local Similarity  44.8%;  Pred. No. 3.93e-17;
  Matches   43;  Conservative   18;  Mismatches  32;  Indels   3;  Gaps   2;
```

```
Db    436 cqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcppgyngkncstp-vsr--c 492
          |:|  |:|| :| :| |||    | ||: | | || |:| |  ||:|  |  |  :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    493 ehnpchngatchersnryvcecargygglncqfllp 528
          |   |:||| | ::::| ||:|  |:|| :|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    7
ID   W00876 standard; Protein; 740 AA.
AC   W00876;
DT   28-APR-1997  (first entry)
DE   C-Delta-1 polypeptide (alternatively spliced variant).
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          184..228
FT                   /label= DSL
FT   domain          229..261
FT                   /label= EGF1
FT   domain          262..292
FT                   /label= EGF2
FT   domain          293..332
FT                   /label= EGF3
FT   domain          333..370
FT                   /label= EGF4
FT   domain          371..409
FT                   /label= EGF5
FT   domain          410..447
FT                   /label= EGF6
FT   domain          448..485
FT                   /label= EGF7
FT   domain          486..523
FT                   /label= EGF8
FT   domain          524..534
FT                   /label= EGF9
FT   domain          555..579
FT                   /label= TM
FT                   /note= "transmembrane domain"
PN   WO9701571-A1.
PD   16-JAN-1997.
PF   28-JUN-1996; U11178.
PR   28-JUN-1995; US-000589.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
DR   WPI; 97-100159/09.
DR   N-PSDB; T58898.
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
PS   Disclosure; Fig 2; 135pp; English.
CC   C-delta-1 polypeptide (W00876) is the chick homologue of Drosophila
CC   Delta, a protein that binds to Notch protein.  Expression of
CC   C-Delta-1 correlates with onset of neurogenesis.  The C-delta-1
CC   amino acid sequence was deduced from a cDNA clone (T58898) obtd.
CC   from chick stage 4-6 embryos.  A shorter version (W58877) of
CC   C-Delta-1, lacking the 12 C-terminal amino acids of the longer
CC   version, was also isolated, and mouse (W11720) and human (W11721-
CC   38) Delta-1 polypeptides have been identified.  Delta-1 proteins
CC   can be used to treat or prevent disorders characterised by
CC   increased Notch activity, such as cervical, breast, lung or colon
CC   cancer, melanoma or seminoma, and nervous system disorders or to
CC   promote tissue regeneration and repair.
SQ   Sequence   740 AA;

  Query Match          28.0%;  Score 325;  DB 21;  Length 740;
  Best Local Similarity  44.8%;  Pred. No. 3.93e-17;
  Matches   43;  Conservative   18;  Mismatches  32;  Indels   3;  Gaps   2;

Db    436 cqcqagftgrhcddnvddcasfpcvnggtcqdgvndysctcppgyngkncstp-vsr--c 492
          |:|  |:|| :| :| |||    | ||: | | || |:| |  ||:|  |  |  :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    493 ehnpchngatchersnryvcecargygglncqfllp 528
          |   |:||| | ::::| ||:|  |:|| :|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    8
ID   W11725 standard; Protein; 660 AA.
AC   W11725;
DT   28-APR-1997  (first entry)
DE   H-Delta-1 polypeptide (reading frame 1 product).
KW   H-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   region          219..221
FT                   /note= "Delta-1 homologous region"
FT   region          245..246
FT                   /note= "Delta-1 homologous region"
FT   region          259..428
FT                   /note= "Delta-1 homologous region"
FT   region          430..434
FT                   /note= "Delta-1 homologous region"
FT   region          594..597
FT                   /note= "Delta-1 homologous region"
FT   region          605..608
FT                   /note= "Delta-1 homologous region"
FT   region          615..617
FT                   /note= "Delta-1 homologous region"
FT   misc_difference 32
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 40
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 41
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 87
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 138
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 139
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 145
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 162
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 187
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 203
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 230
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 249
FT                   /note= "residue corresponds to stop codon in
FT                   H-Delta-1 contig"
FT   misc_difference 429
FT                   /note= "undetermined amino acid residue"
FT   misc_difference 447
```

FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 498
FT /note= "undetermined amino acid residue"
FT misc_difference 513
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 541
FT /note= "undetermined amino acid residue"
FT misc_difference 552
FT /note= "undetermined amino acid residue"
FT misc_difference 556
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 564
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 576
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 580
FT /note= "undetermined amino acid residue"
FT misc_difference 619
FT /note= "undetermined amino acid residue"
FT misc_difference 621
FT /note= "undetermined amino acid residue"
FT misc_difference 626
FT /note= "undetermined amino acid residue"
FT misc_difference 630
FT /note= "undetermined amino acid residue"
FT misc_difference 634
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 639
FT /note= "undetermined amino acid residue"
FT misc_difference 642
FT /note= "undetermined amino acid residue"
FT misc_difference 643
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 644
FT /note= "undetermined amino acid residue"
FT misc_difference 647
FT /note= "residue corresponds to stop codon in
FT H-Delta-1 contig"
FT misc_difference 648
FT /note= "undetermined amino acid residue"
FT misc_difference 651
FT /note= "undetermined amino acid residue"
FT misc_difference 652
FT /note= "undetermined amino acid residue"
PN WO9701571-A1.
PD 16-JAN-1997.
PF 28-JUN-1996; U11178.
PR 28-JUN-1995; US-000589.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique D, Ish-Horowicz D;
PI Lewis J;
DR WPI; 97-100159/09.
DR N-PSDB; T59454.
PT New vertebrate Delta protein, DNA and antibodies - for treating and
PT preventing cancer, nervous system disorders and for tissue
PT regeneration
PS Disclosure; Fig 12B1-6; 135pp; English.
CC Polypeptide sequences (W11725-27) were determined for all 3
CC reading frames of a human H-delta-1 contig sequence (T59454) obtd.
CC from a foetal brain library. Errors in the contig sequence meant
CC that no single reading frame gave the correct sequence for the
CC H-Delta-1 protein. The 3 polypeptide sequences were therefore
CC compared to chick and mouse Delta-1 sequences (see also W11719-20)
CC and regions of homology (see also W11728-38) identified. H-Delta-1
CC is the human homologue of Drosophila Delta, a protein that binds
CC to Notch protein. H-Delta-1 polypeptides can be used to treat
CC disorders of cell fate or differentiation, such as cancer, and
CC nervous system disorders, or to promote tissue regeneration and
CC repair.
SQ Sequence 660 AA;

Query Match 27.8%; Score 323; DB 21; Length 660;
Best Local Similarity 43.8%; Pred. No. 5.73e-17;
Matches 42; Conservative 20; Mismatches 31; Indels 3; Gaps 2;

Db 341 crcqagfsgrhcddnvddcasspcanggtcrdgvndfsctcppgytgrncs-apasr--c 397
Qy 9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db 398 ehapchngatcherghryxcecarsyggpncxfllp 433
Qy 69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104

RESULT 9
ID W11720 standard; Protein; 722 AA.
AC W11720;
DT 28-APR-1997 (first entry)
DE M-Delta-1 polypeptide.
KW M-Delta-1; cell proliferation; nervous system disorder;
KW tissue regeneration; Notch; cervix cancer; breast cancer;
KW lung cancer; colon cancer; melanoma; seminoma;
KW neurogenesis; therapy.
OS Mus sp.
PN WO9701571-A1.
PD 16-JAN-1997.
PF 28-JUN-1996; U11178.
PR 28-JUN-1995; US-000589.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique D, Ish-Horowicz D;
PI Lewis J;
DR WPI; 97-100159/09.
DR N-PSDB; T58899.
PT New vertebrate Delta protein, DNA and antibodies - for treating and
PT preventing cancer, nervous system disorders and for tissue
PT regeneration
PS Claim 4; Fig 8; 135pp; English.
CC M-delta-1 polypeptide (W11720) is the mouse homologue of Drosophila
CC Delta, a protein that binds to Notch protein. It is expressed
CC primarily in presomitic mesoderm, the central and peripheral
CC nervous systems, and kidney. Chick (W11719) and human (W11721-
CC 38) Delta-1 polypeptides have also been identified. Delta-1
CC proteins can be used to treat or prevent disorders characterised by
CC increased Notch activity, such as cervical, breast, lung or colon
CC cancer, melanoma or seminoma, as well as nervous system disorders,
CC and to promote tissue regeneration and repair.
SQ Sequence 722 AA;

Query Match 27.7%; Score 321; DB 21; Length 722;
Best Local Similarity 43.8%; Pred. No. 8.35e-17;
Matches 42; Conservative 19; Mismatches 32; Indels 3; Gaps 2;

Db 428 crcqagfsgrycednvddcasspcanggtcrdsvndfsctcppgytgkncs-apvsr--c 484
Qy 9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db 485 ehapchngatchqrgqrymcecaqgyggpncqfllp 520
Qy 69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104

RESULT 10
ID W18351 standard; protein; 1036 AA.
AC W18351;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;

KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW immunosuppression.
OS Homo sapiens.
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 5; Page 66-71; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1036 AA;

Query Match 25.2%; Score 292; DB 25; Length 1036;
Best Local Similarity 40.0%; Pred. No. 1.92e-14;
Matches 38; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

Db 441 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlcqld-i--dy 497
Qy 8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db 498 cepnpcqngaqcynrasdyfckcpedyegkncshl 532
Qy 68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102

RESULT 11
ID W18352 standard; protein; 1187 AA.
AC W18352;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;
KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW immunosuppression.
OS Homo sapiens.
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 6; Page 71-76; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1187 AA;

Query Match 25.2%; Score 292; DB 25; Length 1187;
Best Local Similarity 40.0%; Pred. No. 1.92e-14;
Matches 38; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

Db 441 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlcqld-i--dy 497
Qy 8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db 498 cepnpcqngaqcynrasdyfckcpedyegkncshl 532
Qy 68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102

RESULT 12
ID W18354 standard; protein; 1218 AA.
AC W18354;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;
KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW immunosuppression.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= Signal
FT Protein 32..1218
FT /label= Differentiation_suppression_protein
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
DR N-PSDB; T70175.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 15; Page 83-91; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1218 AA;

Query Match 25.2%; Score 292; DB 25; Length 1218;
Best Local Similarity 40.0%; Pred. No. 1.92e-14;
Matches 38; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

Db 472 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlcqld-i--dy 528
Qy 8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db 529 cepnpcqngaqcynrasdyfckcpedyegkncshl 563
Qy 68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102

RESULT 13
ID W05833 standard; Protein; 1218 AA.
AC W05833;
DT 28-JAN-1997 (first entry)
DE Human Serrate-1 (HJ1).
KW Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
KW cell fate; central nervous system; cancer; tissue repair; therapy;
KW diagnosis; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..1067
FT /label= Extracellular_domain
FT peptide 14..29
FT /label= Sig_peptide
FT domain 185..229
FT /label= DSL
FT /note= "region of homology with Drosophila Delta
FT and Serrate, predicted to mediate binding
FT with Notch"
FT domain 234..896

```
FT                  /label= ELR
FT                  /note= "epidermal growth factor-like repeat domain"
FT   region         234..264
FT                  /label= ELR1
FT   region         265..299
FT                  /label= ELR2
FT   region         300..339
FT                  /label= ELR3
FT   region         340..377
FT                  /label= ELR4
FT   region         378..415
FT                  /label= ELR5
FT   region         416..453
FT                  /label= ELR6
FT   region         454..490
FT                  /label= ELR7
FT   region         491..528
                    /label= ELR8
     region         529..566
FT                  /label= ELR9
FT   region         567..598
FT                  /label= Partial_ELR
FT   region         599..632
FT                  /label= Partial_ELR
FT   region         633..670
FT                  /label= ELR10
FT   region         671..708
FT                  /label= ELR11
FT   region         709..747
FT                  /label= ELR12
FT   region         748..785
FT                  /label= ELR13
FT   region         786..823
FT                  /label= ELR14
FT   region         824..862
FT                  /label= ELR15
FT   region         863..879
FT                  /label= Partial_ELR
FT   region         880..896
FT                  /label= Partial_ELR
FT   domain         1068..1089
FT                  /label= Transmembrane_domain
FT   domain         1090..1218
FT                  /label= Intracellular_domain
     WO9627610-A1.
     12-SEP-1996.
     07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI   Lewis JH, Mann RS, Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40090.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Claim 4; Page 95-98; 161pp; English.
CC   Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC   for the zygotic neurogenic locus Notch, and are believed to play a
CC   major role in determining cell fates (differentiation) in the
CC   central nervous system. Their amino acid sequences were deduced
CC   from cDNA clones (see also T40090-91) isolated from human foetal
CC   brain cDNA libraries. The proteins, antibodies raised to them,
CC   and encoding nucleic acids can be used in the detection of
CC   Serrate sequences and in the treatment of disorders of cell fate
CC   or differentiation, partic. cancer, nervous system disorders
CC   and in tissue repair or regeneration.
SQ   Sequence   1218 AA;

  Query Match          25.2%;  Score 292;  DB 19;  Length 1218;
  Best Local Similarity  40.0%;  Pred. No. 1.92e-14;
  Matches   38;  Conservative   18;  Mismatches  36;  Indels   3;  Gaps   2;

Db    472 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlcqld-i--dy 528
          || |  ||:||:|  : |:|  : | ||::| :|:| : |||  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    529 cepnpcqngaqcynrasdyfckcpedyegkncshl 563
          ||  ||||| | :::|  | |  : | :|  |
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102


RESULT  14
ID   W44301 standard; Protein; 1218 AA.
AC   W44301;
DT   19-JUN-1998  (first entry)
DE   Human serrate 1.
KW   Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW   leukaemia; endothelial cell; tumour.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Peptide        1..31
FT                  /label= Signal
FT   Protein        32..1218
FT                  /label= Serrate-1
PN   WO9802458-A1.
PD   22-JAN-1998.
PF   11-JUL-1997; J02414.
PR   14-MAY-1997; JP-124063.
PR   16-JUL-1996; JP-186220.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A, Sakano S;
DR   WPI; 98-110528/10.
DR   N-PSDB; V15201.
PT   Human serrate-2 gene expression products - used to regulate stem
PT   cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS   Disclosure; Page 77-86; 103pp; Japanese.
CC   The present sequence represents human serrate 1, from the present
CC   invention which describes human serrate 2. The present invention also
CC   describes a method for the preparation of the polypeptides, and
CC   antibodies binding to the polypeptide and its fragments. The polypeptide
CC   and its fragments expressed by the serrate-2-gene can be used to inhibit
CC   stem (especially blood stem) cell differentiation and to inhibit
CC   endothelial cell growth. They may be incorporated in a cell culture
CC   media for culturing undifferentiated stem cells. They can also be used
CC   for treatment of neoplasms such as leukaemia. The antibodies can be used
CC   for the diagnosis of malignant tumours.
SQ   Sequence   1218 AA;

  Query Match          25.2%;  Score 292;  DB 29;  Length 1218;
  Best Local Similarity  40.0%;  Pred. No. 1.92e-14;
  Matches   38;  Conservative   18;  Mismatches  36;  Indels   3;  Gaps   2;

Db    472 rcicppgyagdhcerdidecasnpclngghcqneinrfqclcptgfsgnlcqld-i--dy 528
          || |  ||:||:|  : |:|  : | ||::| :|:| : |||  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    529 cepnpcqngaqcynrasdyfckcpedyegkncshl 563
          ||  ||||| | :::|  | |  : | :|  |
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102


RESULT  15
ID   W05835 standard; Protein; 1193 AA.
AC   W05835;
DT   28-JAN-1997  (first entry)
DE   Chick Serrate.
KW   C-Serrate; Notch; cell differentiation; cell fate; tissue repair;
KW   central nervous system; cancer; therapy; diagnosis.
OS   Gallus sp.
FH   Key            Location/Qualifiers
FT   domain         1..1041
FT                  /label= Extracellular_domain
FT   peptide        1..5
FT                  /label= Sig_peptide
FT                  /note= "lacks the N-terminal portion owing to
```

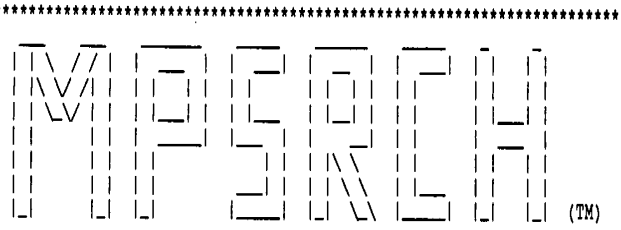
```
FT                truncation of the encoding cDNA clone"
FT   domain       158..203
FT                /label= DSL
FT                /note= "region of homology with Drosophila Delta
FT                and Serrate, predicted to mediate binding
FT                with Notch"
FT   domain       208..837
FT                /label= ELR
FT                /note= "epidermal growth factor-like repeat domain"
FT   region       208..238
FT                /label= ELR1
FT   region       239..274
FT                /label= ELR2
FT   region       275..313
FT                /label= ELR3
FT   region       314..351
FT                /label= ELR4
FT   region       352..390
FT                /label= ELR5
FT   region       391..427
FT                /label= ELR6
FT   region       428..464
FT                /label= ELR7
FT   region       465..502
FT                /label= ELR8
FT   region       503..540
FT                /label= ELR9
FT   region       541..606
FT                /label= ELR10
FT   region       607..644
FT                /label= ELR11
FT   region       655..682
FT                /label= ELR12
FT   region       683..721
FT                /label= ELR13
FT   region       722..759
FT                /label= ELR14
FT   region       760..797
FT                /label= ELR15
FT   region       798..837
FT                /label= ELR16
FT   region       854..911
FT                /label= Cysteine-rich_region
FT   domain       1042..1066
FT                /label= Transmembrane_domain
FT   domain       1067..1193
FT                /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI   Lewis JH, Mann RS, Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40092.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Disclosure; Page 112-115; 161pp; English.
CC   Chicken Serrate (W05835), or C-Serrate, is a ligand for the zygotic
CC   neurogenic locus Notch and is believed to play a major role in
CC   determining cell fates in the central nervous system. Its amino
CC   acid sequence was deduced from a cDNA clone (T40092) obtd. from an
CC   optic explant cDNA library. C-Serrate is expressed in the central
CC   nervous system, cranial placodes, nephric mesoderm, vascular
CC   system, and limb bud mesenchyme.
SQ   Sequence  1193 AA;

 Query Match          25.0%;  Score 290;  DB 19;  Length 1193;
 Best Local Similarity  38.9%;  Pred. No. 2.79e-14;
 Matches   37;  Conservative   18;  Mismatches  37;  Indels   3;  Gaps   2;

Db    446 rcicspgyagdhcekdinecasnpcmngghcqdeingfqclcpagfsgnlcqld-i--dy 502
          || |  ||:||:|  : ::|  : | ||::| ||:|:: |||  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    503 cepnpcqngaqcfnlamdyfcncpedyegkncshl 537
          ||   ||||| | : :    | |   : | :|  |
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102


Search completed: Fri May 28 09:19:22 1999
Job time : 61 secs.
```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:19:40 1999; MasPar time 9.14 Seconds
675.417 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-10
Description: (1-154) from US09191647.pep
Perfect Score: 1160
Sequence: 1 DPLPVHHRCECMLGYTGDNC.....EDNGILLYNGDNDHIAVELY 154

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 39.965; Variance 75.108; scale 0.532

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	462	39.8	1469	2	B36665	slit protein 2 precur	1.19e-72
2	462	39.8	1480	2	A36665	slit protein 1 precur	1.19e-72
3	448	38.6	530	2	A31640	epidermal growth fact	1.10e-69
4	325	28.0	728	2	I50719	C-Delta-1 - chicken	4.53e-44
5	324	27.9	1064	2	A40136	fibropellin Ia - sea	7.25e-44
6	321	27.7	722	2	I48324	DELTA-like 1 - mouse	2.97e-43
7	310	26.7	570	2	A48836	fibropellin C precurs	5.12e-41
8	301	25.9	2703	2	A24420	notch protein - fruit	3.40e-39
9	299	25.8	1203	2	A49175	Motch B protein - mou	8.62e-39
10	299	25.8	2437	2	S42612	transmembrane protein	8.62e-39
11	299	25.8	2471	2	A49128	cell-fate determining	8.62e-39
12	292	25.2	1220	2	A56136	jagged protein precur	2.22e-37
13	288	24.8	2318	2	S45306	notch 3 protein - mou	1.41e-36
14	282	24.3	293	2	B26637	neurogenic repetitive	2.25e-35
15	282	24.3	2139	2	A35672	crumbs protein - frui	2.25e-35
16	282	24.3	2524	2	A35844	Xotch protein - Afric	2.25e-35
17	282	24.3	2531	2	S18188	notch protein homolog	2.25e-35
18	280	24.1	2321	2	S78549	notch3 protein - huma	5.66e-35
19	279	24.1	2531	2	A46019	gene Notch-1 protein	8.96e-35
20	278	24.0	2555	2	A40043	notch protein homolog	1.42e-34
21	265	22.8	1404	2	A36666	serrate protein precu	5.47e-32
22	265	22.8	1408	2	S16148	gene serrate protein	5.47e-32
23	263	22.7	861	2	A48825	Notch homolog Motch p	1.36e-31
24	249	21.5	1429	2	S06434	homeotic protein lin-	7.80e-29
25	242	20.9	833	2	S19087	gene Delta protein pr	1.82e-27
26	240	20.7	387	2	B49175	Motch A protein - mou	4.47e-27
27	240	20.7	832	2	A31246	neurogenic protein De	4.47e-27
28	240	20.7	880	2	S00670	gene Delta protein pr	4.47e-27
29	237	20.4	200	2	A26637	neurogenic repetitive	1.71e-26
30	237	20.4	385	2	S53718	homeotic protein dlk	1.71e-26
31	237	20.4	385	2	A54785	preadipocyte factor 1	1.71e-26
32	228	19.7	1295	2	A32901	glp1 protein precurso	9.47e-25
33	227	19.6	1257	2	S28764	neurocan - rat	1.48e-24
34	225	19.4	260	2	A44549	fetal antigen 1 homeo	3.58e-24
35	225	19.4	1268	2	S52781	neurocan - mouse	3.58e-24
36	223	19.2	383	2	B45484	delta-like dlk homeot	8.68e-24
37	223	19.2	383	2	S53716	homeotic protein dlk	8.68e-24
38	223	19.2	3562	2	A47171	chondroitin sulfate p	8.68e-24
39	222	19.1	259	2	S48713	fetal antigen 1 - hum	1.35e-23
40	221	19.1	4391	2	A38096	perlecan precursor -	2.10e-23
41	219	18.9	102	2	B55885	chondroitin sulfate p	5.07e-23
42	219	18.9	2409	2	A60979	versican precursor -	5.07e-23
43	218	18.8	473	2	A56175	adhesive plaque prote	7.88e-23
44	217	18.7	5147	1	IJFFTM	cadherin-related tumo	1.22e-22
45	215	18.5	2397	2	A55535	versican precursor -	2.94e-22

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           39.8%;  Score 462;  DB 2;  Length 1469;
  Best Local Similarity  42.7%;  Pred. No. 1.19e-72;
  Matches   67;  Conservative   38;  Mismatches  44;  Indels   8;  Gaps   4;

Db   1047 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 1106
          |  |:|  |: | ||::| ||| :|:|||: ||| :| | | | : |:|  ||    |:
Qy      6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db   1107 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 1166
            |::| |:  || :|      |||  :|:| ||: |  || | |::||  ::::::
Qy     62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db   1167 PLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELF 1203
           |: :| ||:|:  |:||:||||:|:|:: |:||||:
Qy    118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


ULT   2
RY            A36665     #type complete
TITLE         slit protein 1 precursor - fruit fly (Drosophila
                melanogaster)
ORGANISM      #formal_name Drosophila melanogaster
DATE          30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                24-Sep-1998
ACCESSIONS    A36665; S13523
REFERENCE     A36665
  #authors    Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                Artavanis-Tsakonas, S.
  #journal    Genes Dev. (1990) 4:2169-2187
  #title      slit: an extracellular protein necessary for development of
                midline glia and commissural axon pathways contains both
                EGF and LRR domains.
  #cross-references MUID:91099665
  #accession  A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene       FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
   66-91          #domain proteoglycan amino-terminal homology #label
                    PAH1\
   101-124        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR1\
   125-148        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR2\
   149-172        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR3\
   173-196        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR4\
   197-220        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR5\
   228-272        #domain proteoglycan carboxyl-terminal homology #label
                    PCS1\
   288-313        #domain proteoglycan amino-terminal homology #label
                    PAH2\
   323-346        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR6\
   347-370        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                    PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                    PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                    PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                    PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR16\
   791-814        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR17\
   815-838        #domain leucine-rich alpha-2-glycoprotein repeat
                    homology #label LR18\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                    PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           39.8%;  Score 462;  DB 2;  Length 1480;
  Best Local Similarity  42.7%;  Pred. No. 1.19e-72;
  Matches   67;  Conservative   38;  Mismatches  44;  Indels   8;  Gaps   4;

Db   1047 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 1106
          |  |:|  |: | ||::| ||| :|:|||: ||| :| | | | : |:|  ||    |:
Qy      6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db   1107 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 1166
            |::| |:  || :|      |||  :|:| ||: |  || | |::||  ::::::
Qy     62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db   1167 PLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELF 1203
```

```
          |: :| ||:|:  |:||:||||:|:|:: |:||||:
Qy    118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT    3
ENTRY           A31640     #type fragment
TITLE           epidermal growth factor-like protein slit - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A31640
REFERENCE       A31640
   #authors     Rothberg, J.M.; Hartley, D.A.; Walther, Z.;
                  Artavanis-Tsakonas, S.
   #journal     Cell (1988) 55:1047-1059
   #title       slit: An EGF-homologous locus of D. melanogaster involved in
                  the development of the embryonic central nervous system.
   #cross-references MUID:89077533
   #accession   A31640
      ##molecule_type DNA
      ##residues      1-530 ##label ROT
      ##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
   #gene        FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
   #introns     470/3
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        growth factor
FEATURE
   148-181           #domain EGF homology #label EGF
SUMMARY         #length 530  #checksum 6330

  Query Match           38.6%;  Score 448;  DB 2;  Length 530;
  Best Local Similarity  41.4%;  Pred. No. 1.10e-69;
  Matches   65;  Conservative   39;  Mismatches  44;  Indels   9;  Gaps   5;

Db    167 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 226
          |  |:|  |: | ||::| ||| :|:|||: ||| :| | | | : |:|  ||    |:
Qy      6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db    227 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 286
            |::| |:  || :|      |||  :|:| ||: |  || | |::||  ::::::
Qy     62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

       287 PLRTRPEANVTI-VFSSGQNGILMYDGQDAHLAVELF 322
           |: :| ||:|: | :: :||||:|:|:: |:||||:
Qy    118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT    4
ENTRY           I50719     #type complete
TITLE           C-Delta-1 - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  14-Aug-1998
ACCESSIONS      I50719
REFERENCE       I50719
   #authors     Henrique, D.; Adam, J.; Myat, A.; Chitnis, A.; Lewis, J.;
                  Ish-Horowicz, D.
   #journal     Nature (1995) 375:787-790
   #title       Expression of a Delta homologue in prospective neurons in the
                  chick.
   #cross-references MUID:95319507
   #accession   I50719
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-728 ##label HEN
      ##cross-references EMBL:U26590; NID:g882411; PID:g882412
CLASSIFICATION  #superfamily EGF homology
FEATURE
   454-485           #domain EGF homology #label EGF
SUMMARY         #length 728  #molecular-weight 79861  #checksum 1765

  Query Match           28.0%;  Score 325;  DB 2;  Length 728;
  Best Local Similarity  44.8%;  Pred. No. 4.53e-44;
  Matches   43;  Conservative   18;  Mismatches  32;  Indels   3;  Gaps   2;

Db    436 CQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCPPGYNGKNCSTP-VSR--C 492
          |:|  |:|| :| :| |||    | ||: | | || |:| |  ||:|  |  |  :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    493 EHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
          |   |:||| | ::::| ||:|  |:|| :|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    5
ENTRY           A40136     #type complete
TITLE           fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
ALTERNATE_NAMES epidermal growth factor homolog precursor
CONTAINS        alternatively spliced fibropellin Ib (EGFI)
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            13-May-1992 #sequence_revision 17-Sep-1997 #text_change
                  07-Aug-1998
ACCESSIONS      A40136; B40136; C40136; A29316; A43131
REFERENCE       A40136
   #authors     Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff,
                  R.A.
   #journal     J. Mol. Evol. (1989) 29:314-327
   #title       Structural analysis of the uEGF gene in the sea urchin
                  Strongylocentrotus purpuratus reveals more similarity to
                  vertebrate than to invertebrate genes with EGF-like
                  repeats.
   #cross-references MUID:90112459
   #accession   A40136
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-114 ##label DEL
      ##cross-references GB:X17530; NID:g10225; PID:g667061
   #accession   B40136
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues      181-251,329-370,'R',372-408,'RA',411-441 ##label DE2
   #accession   C40136
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues      'K',747-821,898-978 ##label DE3
REFERENCE       A29316
   #authors     Hursh, D.A.; Andrews, M.E.; Raff, R.A.
   #journal     Science (1987) 237:1487-1490
   #title       A sea urchin gene encodes a polypeptide homologous to
                  epidermal growth factor.
   #cross-references MUID:87319677
   #accession   A29316
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      'S',280-481,786-1064 ##label HUR
      ##cross-references GB:M17421; NID:g161474; PID:g552260
REFERENCE       A43131
   #authors     Hunt, L.T.; Barker, W.C.
   #journal     FASEB J. (1989) 3:1760-1764
   #title       Avidin-like domain in an epidermal growth factor homolog from
                  a sea urchin.
   #cross-references MUID:89196806
   #contents    annotation
COMMENT         EGF homology repeats 10-17 are spliced out in the short form
                  (fibropellin Ib).
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-19              #domain signal sequence #status predicted #label SIG\
   20-1064           #product fibropellin I #status predicted #label FIB\
   23-54             #domain EGF homology #label EG01\
   57-175            #domain C1r/C1s repeat homology #label CSR\
   180-211           #domain EGF homology #label EG02\
```

```
   218-249          #domain EGF homology #label EG03\
   256-287          #domain EGF homology #label EG04\
   294-325          #domain EGF homology #label EG05\
   332-363          #domain EGF homology #label EG06\
   370-401          #domain EGF homology #label EG07\
   408-439          #domain EGF homology #label EG08\
   446-477          #domain EGF homology #label EG09\
   484-515          #domain EGF homology #label EG10\
   522-553          #domain EGF homology #label EG11\
   560-591          #domain EGF homology #label EG12\
   598-629          #domain EGF homology #label EG13\
   636-667          #domain EGF homology #label EG14\
   674-705          #domain EGF homology #label EG15\
   712-743          #domain EGF homology #label EG16\
   750-781          #domain EGF homology #label EG17\
   788-819          #domain EGF homology #label EG18\
   826-857          #domain EGF homology #label EG19\
   864-895          #domain EGF homology #label EG20\
   902-933          #domain EGF homology #label EG21\
   936-1064         #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439,
   446-457,451-466,
   468-477,484-495  #disulfide_bonds #status predicted\
   489-504,506-515,
   522-533,527-542,
   544-553,560-571,
   565-580,582-591,
   598-609,603-618,
   620-629,636-647,
   641-656,658-667,
   674-685,679-694,
   696-705,712-723,
   717-732,734-743,
   750-761,755-770,
   772-781,788-799,
   793-808,810-819,
   826-837,831-846,
   848-857,864-875,
   869-884,886-895,
   902-913,907-922,
   924-933          #disulfide_bonds #status predicted
SUMMARY          #length 1064  #molecular-weight 112072  #checksum 303

  Query Match           27.9%;  Score 324;  DB 2;  Length 1064;
  Best Local Similarity  46.7%;  Pred. No. 7.25e-44;
  Matches   43;  Conservative   16;  Mismatches  30;  Indels   3;  Gaps   2;

Db    808 CACVPGFTGSNCETNIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCEIS-L--DAC 864
          | |: |:|| ||  | |:|    | ||: ||| ||:: | |   ||| |||:     :|
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    865 RSMPCQNGATCVNVGADYVCECVPGYAGQNCE 896
            :  ||||| ||: |:  ||:|:||::| :||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    6
ENTRY           I48324     #type complete
TITLE           DELTA-like 1 - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                  28-Feb-1997
ACCESSIONS      I48324
REFERENCE       I48324
  #authors      Bettenhausen, B.; de Angelis, M.H.; Simon, D.; Guenet, J.L.;
                  Gossler, A.
  #journal      Development (1995) 121:2407-2418
  #title        Transient and restricted expression during mouse
                  embryogenesis of Dll1, a murine gene closely related to
                  Drosophila Delta.
  #cross-references MUID:95401858
  #accession    I48324
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues     1-722 ##label RES
    ##cross-references EMBL:X80903; NID:g806569; PID:g806570
GENETICS
  #gene         Dll1
SUMMARY         #length 722  #molecular-weight 78448  #checksum 1452

  Query Match           27.7%;  Score 321;  DB 2;  Length 722;
  Best Local Similarity  43.8%;  Pred. No. 2.97e-43;
  Matches   42;  Conservative   19;  Mismatches  32;  Indels   3;  Gaps   2;

Db    428 CRCQAGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGKNCS-APVSR--C 484
          | |  |::|  | :| |||    | ||: | | || ::| |  ||:|  |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    485 EHAPCHNGATCHQRGQRYMCECAQGYGGPNCQFLLP 520
          | : |:||| | ::| | :|:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    7
ENTRY           A48836     #type complete
TITLE           fibropellin C precursor - sea urchin (Strongylocentrotus
                  purpuratus)
ALTERNATE_NAMES EGF repeat-containing protein; epidermal growth
                  factor-related protein 3; fibropellin III
ORGANISM        #formal_name Strongylocentrotus purpuratus #common_name
                  purple urchin
DATE            01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  07-Aug-1998
ACCESSIONS      A48836
REFERENCE       A48836
  #authors      Bisgrove, B.W.; Raff, R.A.
  #journal      Dev. Biol. (1993) 157:526-538
  #title        The SpEGF III gene encodes a member of the fibropellins: EGF
                  repeat-containing proteins that form the apical lamina of
                  the sea urchin embryo.
  #cross-references MUID:93273088
  #accession    A48836
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-570 ##label BIS
    ##cross-references GB:L07045; NID:g310659; PID:g310660
    ##note         sequence extracted from NCBI backbone (NCBIN:132724,
                    NCBIP:132725)
CLASSIFICATION  #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
  1-18            #domain signal sequence #status predicted #label SIG\
  19-570          #product fibropellin C #status predicted #label FIB\
  19-54           #domain EGF homology #label EGF1\
  57-175          #domain C1r/C1s repeat homology #label C1R\
  176-211         #domain EGF homology #label EGF2\
  214-249         #domain EGF homology #label EGF3\
  252-287         #domain EGF homology #label EGF4\
  290-325         #domain EGF homology #label EGF5\
  328-363         #domain EGF homology #label EGF6\
  366-401         #domain EGF homology #label EGF7\
  404-439         #domain EGF homology #label EGF8\
  442-570         #region avidin-like\
  23-34,28-43,45-54,
  62-88,180-191,
  185-200,202-211,
```

```
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439    #disulfide_bonds #status predicted
SUMMARY          #length 570  #molecular-weight 61115  #checksum 5567

  Query Match           26.7%;  Score 310;  DB 2;  Length 570;
  Best Local Similarity  42.4%;  Pred. No. 5.12e-41;
  Matches    39;  Conservative   19;  Mismatches  31;  Indels   3;  Gaps   2;

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYVCQCLPNYTGTHCEIS-L--DAC 370
          |:|  |:|| ||  | ::|    | ||: |:| |::| | |:  |:|  |||:     :|
        9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    371 ASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402
           :  ||||: | : |:  ||:||||: | :||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    8
ENTRY            A24420     #type complete
TITLE            notch protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  neurogenic repetitive locus protein
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
                   07-Aug-1998
ACCESSIONS       A24420; A24768; S09358; A05267
REFERENCE        A24420
   #authors      Kidd, S.; Kelley, M.R.; Young, M.W.
   #journal      Mol. Cell. Biol. (1986) 6:3094-3108
   #cross-references MUID:87064624
   #accession    A24420
      ##molecule_type DNA
      ##residues      1-2703 ##label KID
      ##cross-references GB:K03508; NID:g157991; PID:g157993
REFERENCE        A24768
   #authors      Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
   #journal      Cell (1985) 43:567-581
   #cross-references MUID:86079539
   #accession    A24768
      ##molecule_type mRNA
      ##residues      1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',
                      268-872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',
                      2258-2264,'V',2266-2406,'R',2408-2444,'L',2446-2703
                      ##label WHA1
      ##note          the authors translated the codon ATC for residue 49 as
                      Thr, ATT for residue 2044 as Arg, GTA for residue 2265
                      as Ala, CGC for residue 2407 as His, and CTT for
                      residue 2445 as Arg
REFERENCE        S09358
   #authors      Tautz, D.
   #journal      Nucleic Acids Res. (1989) 17:6463-6471
   #title        Hypervariability of simple sequences as a general source for
                   polymorphic DNA markers.
   #cross-references MUID:89385974
   #accession    S09358
      ##molecule_type DNA
      ##residues      2505-2551,'QQQQ',2552-2576,'E',2578-2604 ##label TAU
REFERENCE        A05267
   #authors      Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.;
                   Artavanis-Tsakonas, S.
   #journal      Cell (1985) 40:55-62
   #title        opa: a novel family of transcribed repeats shared by the
                   Notch locus and other developmentally regulated loci in D.
                   melanogaster.
   #cross-references MUID:85099329
   #accession    A05267
      ##molecule_type DNA
      ##residues      2504-2576,'E',2578-2611 ##label WHA2
GENETICS
   #gene         notch; opa
      ##cross-references FlyBase:FBgn0004647
   #map_position 8.96-9.36
   #introns      53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
CLASSIFICATION   #superfamily notch protein; ankyrin repeat homology; EGF
                   homology
KEYWORDS         differentiation; tandem repeat; transmembrane protein
FEATURE
   27-43             #domain transmembrane #status predicted #label TMM1\
   568-599           #domain EGF homology #label EGF\
   1746-1762         #domain transmembrane #status predicted #label TMM2\
   1950-1982         #domain ankyrin repeat homology #label AN1\
   1983-2015         #domain ankyrin repeat homology #label AN2\
   1988-2004         #domain transmembrane #status predicted #label TMM3\
   2017-2049         #domain ankyrin repeat homology #label AN3\
   2050-2082         #domain ankyrin repeat homology #label AN4\
   2083-2115         #domain ankyrin repeat homology #label AN5\
   2538-2568         #region glutamine-rich\
   2538-2568         #domain neurogenic repetitive element #status predicted
                       #label OPA
SUMMARY          #length 2703  #molecular-weight 288876  #checksum 6404

  Query Match           25.9%;  Score 301;  DB 2;  Length 2703;
  Best Local Similarity  40.7%;  Pred. No. 3.40e-39;
  Matches    37;  Conservative   19;  Mismatches  32;  Indels   3;  Gaps   2;

Db   1008 CTCPLGFSGINCQTNDEDCTESSCLNGGSCIDGINGYNCSCLAGYSGANCQYKLN-K--C 1064
          | | ||::| ||  |::|| :  | ||: |:| :|:| | |: ||||  |:     :  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db   1065 DSNPCLNGATCHEQNNEYTCHCPSGFTGKQC 1095
           ::  | ||| | :| :   |:| :|| | :|
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPEC 99


RESULT    9
ENTRY            A49175     #type fragment
TITLE            Motch B protein - mouse (fragment)
ALTERNATE_NAMES  Notch homolog
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                   14-Aug-1998
ACCESSIONS       A49175; PH1570; S32113
REFERENCE        A49175
   #authors      Lardelli, M.; Lendahl, U.
   #journal      Exp. Cell Res. (1993) 204:364-372
   #title        Motch A and Motch B--two mouse Notch homologues coexpressed
                   in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession    A49175
      ##status        preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-1203 ##label LAR
      ##cross-references EMBL:X68279; NID:g287989; PID:g287990
      ##experimental_source embryo
      ##note          sequence extracted from NCBI backbone (NCBIP:126158)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermaland neural fate for cells in the early
             embryo.
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   560-591           #domain EGF homology #label EGF
SUMMARY          #length 1203  #checksum 910

  Query Match           25.8%;  Score 299;  DB 2;  Length 1203;
  Best Local Similarity  38.7%;  Pred. No. 8.62e-39;
  Matches    36;  Conservative   20;  Mismatches  34;  Indels   3;  Gaps   2;

Db    124 HCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFKGVHCELE-V--NE 180
```

```
          :|||: ||:|  |  : ::|    ||| | |:| : :::|||: |: |  ||:     :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    181 CQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQ 213
          |::  | | : |||  :|  | | ||| || |:
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   10
ENTRY            S42612     #type complete
TITLE            transmembrane protein precursor - zebra fish
ORGANISM         #formal_name Brachydanio rerio #common_name zebra fish
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   10-Jul-1998
ACCESSIONS       S42612
REFERENCE        S42612
   #authors      Bierkamp, C.; Campos-Ortega, J.A.
   #journal      Mech. Dev. (1993) 43:87-100
   #title        A zebrafish homologue of the Drosophila neurogenic gene Notch
                   and its pattern of transcription during early
                   embryogenesis.
   #cross-references MUID:94128602
   #accession    S42612
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-2437 ##label BIE
      ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1915-1947          #domain ankyrin repeat homology #label AN1\
   1948-1980          #domain ankyrin repeat homology #label AN2\
   1982-2014          #domain ankyrin repeat homology #label AN3\
   2015-2047          #domain ankyrin repeat homology #label AN4\
   2048-2080          #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           25.8%;  Score 299;  DB 2;  Length 2437;
  Best Local Similarity  44.1%;  Pred. No. 8.62e-39;
  Matches   41;  Conservative   14;  Mismatches  34;  Indels   4;  Gaps   3;

Db    474 HCICMPGYEGVFCQINSDDCASQPCLNG-KCIDKINSFHCECPKGFSGSLCQVD-V--DE 529
          :| || || |  |  | |||  : | ||  |:| :||: | |  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    530 CASTPCKNGAKCTDGPNKYTCECTPGFSGIHCE 562
          | :| | |||:| |  ::  |:| |||:|  ||
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   11
ENTRY            A49128     #type complete
TITLE            cell-fate determining gene Notch2 protein - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                   14-Aug-1998
ACCESSIONS       A49128
REFERENCE        A49128
   #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal      Development (1992) 116:931-941
   #title        Notch2: a second mammalian Notch gene.
   #cross-references MUID:93202015
   #accession    A49128
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues      1-2471 ##label WEI
      ##experimental_source Schwann cell
      ##note          sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   1029-1060          #domain EGF homology #label EGF\
   1876-1908          #domain ankyrin repeat homology #label AN1\
   1909-1941          #domain ankyrin repeat homology #label AN2\
   1943-1975          #domain ankyrin repeat homology #label AN3\
   1976-2008          #domain ankyrin repeat homology #label AN4\
   2009-2041          #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           25.8%;  Score 299;  DB 2;  Length 2471;
  Best Local Similarity  38.7%;  Pred. No. 8.62e-39;
  Matches   36;  Conservative   20;  Mismatches  34;  Indels   3;  Gaps   2;

Db    441 HCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFKGVHCELE-V--NE 497
          :|||: ||:|  |  : ::|    ||| | |:| : :::|||: |: |  ||:     :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    498 CQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQ 530
          |::  | | : |||  :|  | | ||| || |:
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   12
ENTRY            A56136     #type complete
TITLE            jagged protein precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change
                   11-Aug-1995
ACCESSIONS       A56136
REFERENCE        A56136
   #authors      Lindsell, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.
   #journal      Cell (1995) 80:909-917
   #title        Jagged: a mammalian ligand that activates Notch1.
   #cross-references MUID:95211842
   #accession    A56136
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-1220 ##label LIN
      ##cross-references GB:L38483
SUMMARY          #length 1220  #molecular-weight 134528  #checksum 2746

  Query Match           25.2%;  Score 292;  DB 2;  Length 1220;
  Best Local Similarity  40.0%;  Pred. No. 2.22e-37;
  Matches   38;  Conservative   18;  Mismatches  36;  Indels   3;  Gaps   2;

Db    473 RCICPPGYAGDHCERDIDECASNPCLNGGHCQNEINRFQCLCPTGFSGNLCQLD-I--DY 529
          || |  ||:||:|  : |:|  : | ||::| :|:| : |||  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    530 CEPNPCQNGAQCYNRASDYFCKCPEDYEGKNCSHL 564
          ||   ||||| | :::|   | |   : | :|  |
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKL 102


RESULT   13
ENTRY            S45306     #type complete
TITLE            notch 3 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   10-Jul-1998
ACCESSIONS       S45306
REFERENCE        S45306
   #authors      Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal      Mech. Dev. (1994) 46:123-136
   #title        The novel Notch homologue mouse Notch 3 lacks specific
                   epidermal growth factor-repeats and is expressed in
                   proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession    S45306
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
```

```
   1839-1871          #domain ankyrin repeat homology #label AN1\
   1872-1904          #domain ankyrin repeat homology #label AN2\
   1906-1938          #domain ankyrin repeat homology #label AN3\
   1939-1971          #domain ankyrin repeat homology #label AN4\
   1972-2004          #domain ankyrin repeat homology #label AN5
SUMMARY           #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           24.8%;  Score 288;  DB 2;  Length 2318;
  Best Local Similarity  41.9%;  Pred. No. 1.41e-36;
  Matches    39;  Conservative   16;  Mismatches  35;  Indels   3;  Gaps   2;

Db    456 CICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMCQLD-V--DEC 512
          | || |:||  |  : |:|    | ||: | | ||:::| |  |:|| :|::      |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    513 ASTPCRNGAKCVDQPDGYECRCAEGFEGTLCER 545
          :| |:|||:||||     |:|  || |  ||:
      69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEK 101


RESULT   14
ENTRY             B26637     #type fragment
TITLE             neurogenic repetitive locus 95F protein - fruit fly
                    (Drosophila melanogaster) (fragment)
ORGANISM          #formal_name Drosophila melanogaster
DATE              16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
                    14-Aug-1998
ACCESSIONS        B26637
REFERENCE         A91081
   #authors       Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigel,
                    D.; Vaessin, H.; Campos-Ortega, J.A.
   #journal       EMBO J. (1987) 6:761-766
   #title         EGF homologous sequences encoded in the genome of Drosophila
                    melanogaster, and their relation to neurogenic genes.
   #cross-references MUID:87218537
   #accession     B26637
      ##molecule_type mRNA
      ##residues      1-293 ##label KNU
      ##cross-references GB:X05144; NID:g7519; PID:g929536
GENETICS
   #gene          FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION    #superfamily EGF homology
KEYWORDS          transmembrane protein
FEATURE
   216-252            #domain EGF homology #label EGF
MARY              #length 293  #checksum 3413

  Query Match           24.3%;  Score 282;  DB 2;  Length 293;
  Best Local Similarity  39.6%;  Pred. No. 2.25e-35;
  Matches    38;  Conservative   19;  Mismatches  34;  Indels   5;  Gaps   5;

Db    159 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTCEN 218
          |:|  |: |::| :| |:| |: |:||: | | : || | | | | |  |::    :
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEI-PP-A-PR 65

Db    219 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 253
          :| :|: |||| |    |:   | |:||| || |:
Qy     66 SSC-EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   15
ENTRY             A35672     #type complete
TITLE             crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM          #formal_name Drosophila melanogaster
DATE              21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                    14-Aug-1998
ACCESSIONS        A35672
REFERENCE         A35672
   #authors       Tepass, U.; Theres, C.; Knust, E.
   #journal       Cell (1990) 61:787-799
   #title         crumbs encodes an EGF-like protein expressed on apical
                    membranes of Drosophila epithelial cells and required for
                    organization of epithelia.
   #cross-references MUID:90263104
   #accession     A35672
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-2139 ##label TEP
      ##cross-references GB:M33753
      ##note          the authors translated the codon GGC for residue 1928 as
                       Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene          FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION    #superfamily EGF homology
KEYWORDS          transmembrane protein
FEATURE
   691-722            #domain EGF homology #label EGF
SUMMARY           #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           24.3%;  Score 282;  DB 2;  Length 2139;
  Best Local Similarity  39.6%;  Pred. No. 2.25e-35;
  Matches    38;  Conservative   19;  Mismatches  34;  Indels   5;  Gaps   5;

Db   1821 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTCEN 1880
          |:|  |: |::| :| |:| |: |:||: | | : || | | | | |  |::    :
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEI-PP-A-PR 65

Db   1881 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 1915
          :| :|: |||| |    |:   | |:||| || |:
Qy     66 SSC-EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


Search completed: Fri May 28 09:20:00 1999
Job time : 20 secs.
```

MASPAR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:20:17 1999; MasPar time 6.36 Seconds
684.284 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-10
Description: (1-154) from US09191647.pep
Perfect Score: 1160
Sequence: 1 DPLPVHHRCECMLGYTGDNC.....EDNGILLYNGDNDHIAVELY 154

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 40.992; Variance 67.443; scale 0.608

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	462	39.8	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	7.20e-83
2	324	27.9	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	4.14e-50
3	321	27.7	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	2.05e-49
4	321	27.7	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	2.05e-49
5	315	27.2	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	5.03e-48
6	310	26.7	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	7.18e-47
7	301	25.9	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	8.47e-45
8	299	25.8	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	2.44e-44
9	288	24.8	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	8.01e-42
10	282	24.3	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	1.86e-40
11	282	24.3	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	1.86e-40
12	282	24.3	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	1.86e-40
13	279	24.1	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	8.95e-40
14	278	24.0	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	1.51e-39
15	272	23.4	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	3.45e-38
16	265	22.8	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	1.31e-36
17	249	21.5	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	4.99e-33
18	240	20.7	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	4.94e-31
19	237	20.4	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	2.27e-30
20	228	19.7	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	2.15e-28
21	227	19.6	1257	1	PGCN_RAT	NEUROCAN CORE PROTEIN	3.56e-28
22	225	19.4	1268	1	PGCN_MOUSE	NEUROCAN CORE PROTEIN	9.75e-28
23	223	19.2	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	2.66e-27
24	223	19.2	3562	1	PGCV_CHICK	VERSICAN CORE PROTEIN	2.66e-27
25	221	19.1	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	7.24e-27
26	219	18.9	3396	1	PGCV_HUMAN	VERSICAN CORE PROTEIN	1.97e-26
27	217	18.7	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	5.34e-26
28	215	18.5	3358	1	PGCV_MOUSE	VERSICAN CORE PROTEIN	1.45e-25
29	212	18.3	862	1	PGCV_MACNE	VERSICAN CORE PROTEIN	6.42e-25
30	205	17.7	4544	1	LRP1_HUMAN	LOW-DENSITY LIPOPROTEI	2.04e-23
31	200	17.2	432	1	NEL2_RAT	NEL-LIKE PROTEIN (FRAG	2.38e-22
32	199	17.2	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	3.88e-22
33	192	16.6	428	1	NEL2_HUMAN	NEL-LIKE PROTEIN (FRAG	1.17e-20
34	193	16.6	1376	1	NID2_HUMAN	NIDOGEN-2 PRECURSOR (N	7.20e-21
35	184	15.9	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	5.53e-19
36	183	15.8	515	1	APX1_CAEEL	APX-1 PROTEIN PRECURSO	8.92e-19
37	183	15.8	1959	1	AGRI_RAT	AGRIN PRECURSOR.	8.92e-19
38	181	15.6	2109	1	PGCA_CHICK	AGGRECAN CORE PROTEIN	2.32e-18
39	174	15.0	816	1	NEL_HUMAN	NEL PROTEIN PRECURSOR	6.43e-17
40	174	15.0	4543	1	LRP1_CHICK	LOW-DENSITY LIPOPROTEI	6.43e-17
41	173	14.9	816	1	NEL_CHICK	NEL PROTEIN PRECURSOR	1.03e-16
42	172	14.8	564	1	PGCA_CANFA	AGGRECAN CORE PROTEIN	1.65e-16
43	172	14.8	816	1	NEL_MOUSE	NEL PROTEIN PRECURSOR	1.65e-16
44	172	14.8	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	1.65e-16
45	171	14.7	2871	1	FBN1_HUMAN	FIBRILLIN 1 PRECURSOR.	2.64e-16

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE  1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          39.8%;  Score 462;  DB 1;  Length 1480;
  Best Local Similarity  42.7%;  Pred. No. 7.20e-83;
  Matches   67;  Conservative   38;  Mismatches  44;  Indels   8;  Gaps   4;

Db  1047 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 1106
              |  |:|  |: | ||::| ||| :|:|||: ||| :| | | | : |:|  ||    |:
Qy     6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db  1107 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 1166
               |::| |:  || :|       |||  :|:| ||: |  || | |::||  ::::::
Qy    62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db  1167 PLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELF 1203
              |: :| ||:|:  |:||:||||:|:|:: |:||||:
Qy   118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT   2
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
```

```
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1    19       POTENTIAL.
FT   CHAIN       20  1064       FIBROPELLIN I.
FT   DOMAIN      20    55       EGF-LIKE 1.
FT   DOMAIN      62   175       CUB.
FT   DOMAIN     176   212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214   250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252   288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290   326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328   364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366   402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404   440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442   478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480   516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     518   554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     556   592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     594   630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632   668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     670   706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     708   744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     746   782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     784   820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     822   858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     860   896       EGF-LIKE 20.
FT   DOMAIN     898   934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     936  1064       AVIDIN-LIKE.
FT   DISULFID    23    34       BY SIMILARITY.
FT   DISULFID    28    43       BY SIMILARITY.
FT   DISULFID    45    54       BY SIMILARITY.
FT   DISULFID   180   191       BY SIMILARITY.
FT   DISULFID   185   200       BY SIMILARITY.
FT   DISULFID   202   211       BY SIMILARITY.
FT   DISULFID   218   229       BY SIMILARITY.
FT   DISULFID   223   238       BY SIMILARITY.
FT   DISULFID   240   249       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   261   276       BY SIMILARITY.
FT   DISULFID   278   287       BY SIMILARITY.
FT   DISULFID   294   305       BY SIMILARITY.
FT   DISULFID   299   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   457       BY SIMILARITY.
FT   DISULFID   451   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   DISULFID   522   533       BY SIMILARITY.
FT   DISULFID   527   542       BY SIMILARITY.
FT   DISULFID   544   553       BY SIMILARITY.
FT   DISULFID   560   571       BY SIMILARITY.
FT   DISULFID   565   580       BY SIMILARITY.
FT   DISULFID   582   591       BY SIMILARITY.
FT   DISULFID   598   609       BY SIMILARITY.
FT   DISULFID   603   618       BY SIMILARITY.
FT   DISULFID   620   629       BY SIMILARITY.
FT   DISULFID   636   647       BY SIMILARITY.
FT   DISULFID   641   656       BY SIMILARITY.
FT   DISULFID   658   667       BY SIMILARITY.
FT   DISULFID   674   685       BY SIMILARITY.
FT   DISULFID   679   694       BY SIMILARITY.
FT   DISULFID   696   705       BY SIMILARITY.
FT   DISULFID   712   723       BY SIMILARITY.
FT   DISULFID   717   732       BY SIMILARITY.
FT   DISULFID   734   743       BY SIMILARITY.
FT   DISULFID   750   761       BY SIMILARITY.
FT   DISULFID   755   770       BY SIMILARITY.
FT   DISULFID   772   781       BY SIMILARITY.
FT   DISULFID   788   799       BY SIMILARITY.
FT   DISULFID   793   808       BY SIMILARITY.
FT   DISULFID   810   819       BY SIMILARITY.
FT   DISULFID   826   837       BY SIMILARITY.
FT   DISULFID   831   846       BY SIMILARITY.
FT   DISULFID   848   857       BY SIMILARITY.
FT   DISULFID   864   875       BY SIMILARITY.
FT   DISULFID   869   884       BY SIMILARITY.
FT   DISULFID   886   895       BY SIMILARITY.
FT   DISULFID   902   913       BY SIMILARITY.
FT   DISULFID   907   922       BY SIMILARITY.
FT   DISULFID   924   933       BY SIMILARITY.
FT   VARSPLIC   477   780       MISSING (IN FORM IB).
FT   CARBOHYD    30    30       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   851   851       POTENTIAL.
FT   CONFLICT   279   279       L -> S (IN REF. 2).
```

```
SQ   SEQUENCE   1064 AA;  112072 MW;  FBD10D48 CRC32;

  Query Match           27.9%;  Score 324;  DB 1;  Length 1064;
  Best Local Similarity  46.7%;  Pred. No. 4.14e-50;
  Matches   43;  Conservative   16;  Mismatches  30;  Indels   3;  Gaps   2;

Db    808 CACVPGFTGSNCETNIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCEIS-L--DAC 864
          | |: |:|| ||  | |:|    | ||: ||| ||:: | |   ||| |||:    :|
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    865 RSMPCQNGATCVNVGADYVCECVPGYAGQNCE 896
            :  ||||| ||: |:  ||:|:||::| :||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    3
ID   DLL1_MOUSE     STANDARD;      PRT;   722 AA.
AC   Q61483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C X C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95401858.
RA   BETTENHAUSEN B., DE ANGELIS M.H., SIMON D., GUENET J.-L., GOSSLER A.;
RT   "Transient and restricted expression during mouse embryogenesis of
RT   Dll1, a murine gene closely related to Drosophila Delta.";
RL   DEVELOPMENT 121:2407-2418(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN THE PARAXIAL
CC       MESODERM AND NERVOUS SYSTEM. EXPRESSED AT HIGH LEVELS IN ADULT
CC       HEART AND AT LOWER LEVELS, IN ADULT LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED UNTIL DAY 15 IN THE EMBRYO.
CC       EXPRESSION THEN DECREASES AND INCREASES AGAIN IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80903; G806570; -.
DR   MGD; MGI:104659; DLL1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    722       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546   568       POTENTIAL.
FT   DOMAIN      569   722       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      225   253       EGF-LIKE 1.
FT   DOMAIN      256   284       EGF-LIKE 2.
FT   DOMAIN      291   324       EGF-LIKE 3.
FT   DOMAIN      331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN      369   401       EGF-LIKE 5.
FT   DOMAIN      408   439       EGF-LIKE 6.
FT   DOMAIN      446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      484   515       EGF-LIKE 8.
FT   DISULFID    225   236       BY SIMILARITY.
FT   DISULFID    229   242       BY SIMILARITY.
FT   DISULFID    244   253       BY SIMILARITY.
FT   DISULFID    256   267       BY SIMILARITY.
FT   DISULFID    262   273       BY SIMILARITY.
FT   DISULFID    275   284       BY SIMILARITY.
FT   DISULFID    291   303       BY SIMILARITY.
FT   DISULFID    297   313       BY SIMILARITY.
FT   DISULFID    315   324       BY SIMILARITY.
FT   DISULFID    331   342       BY SIMILARITY.
FT   DISULFID    336   351       BY SIMILARITY.
FT   DISULFID    353   362       BY SIMILARITY.
FT   DISULFID    369   380       BY SIMILARITY.
FT   DISULFID    374   390       BY SIMILARITY.
FT   DISULFID    392   401       BY SIMILARITY.
FT   DISULFID    408   419       BY SIMILARITY.
FT   DISULFID    413   428       BY SIMILARITY.
FT   DISULFID    430   439       BY SIMILARITY.
FT   DISULFID    446   466       BY SIMILARITY.
FT   DISULFID    468   477       BY SIMILARITY.
FT   DISULFID    484   495       BY SIMILARITY.
FT   DISULFID    489   504       BY SIMILARITY.
FT   DISULFID    506   515       BY SIMILARITY.
FT   CARBOHYD    476   476       POTENTIAL.
SQ   SEQUENCE   722 AA;  78448 MW;  5A647702 CRC32;

  Query Match           27.7%;  Score 321;  DB 1;  Length 722;
  Best Local Similarity  43.8%;  Pred. No. 2.05e-49;
  Matches   42;  Conservative   19;  Mismatches  32;  Indels   3;  Gaps   2;

Db    428 CRCQAGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGKNCS-APVSR--C 484
          | |  |::|  | :| |||    | ||: | | || ::| |  ||:|  |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    485 EHAPCHNGATCHQRGQRYMCECAQGYGGPNCQFLLP 520
          | : |:||| | ::| |.:|:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    4
ID   DLL1_HUMAN     STANDARD;      PRT;   723 AA.
AC   O00548;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MANN R.S., GRAY G.E., HENRIQUE D., ISH-HOROWICZ D.,
RA   ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; AF003522; G2197069; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   723       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   723       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     226   254       EGF-LIKE 1.
FT   DOMAIN     257   285       EGF-LIKE 2.
FT   DOMAIN     292   325       EGF-LIKE 3.
FT   DOMAIN     332   363       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     370   402       EGF-LIKE 5.
FT   DOMAIN     409   440       EGF-LIKE 6.
FT   DOMAIN     447   478       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     485   516       EGF-LIKE 8.
FT   DISULFID   226   237       BY SIMILARITY.
FT   DISULFID   230   243       BY SIMILARITY.
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   257   268       BY SIMILARITY.
FT   DISULFID   263   274       BY SIMILARITY.
FT   DISULFID   276   285       BY SIMILARITY.
FT   DISULFID   292   304       BY SIMILARITY.
FT   DISULFID   298   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   391       BY SIMILARITY.
FT   DISULFID   393   402       BY SIMILARITY.
FT   DISULFID   409   420       BY SIMILARITY.
FT   DISULFID   414   429       BY SIMILARITY.
FT   DISULFID   431   440       BY SIMILARITY.
FT   DISULFID   447   467       BY SIMILARITY.
FT   DISULFID   469   478       BY SIMILARITY.
FT   DISULFID   485   496       BY SIMILARITY.
FT   DISULFID   490   505       BY SIMILARITY.
FT   DISULFID   507   516       BY SIMILARITY.
FT   CARBOHYD   477   477       POTENTIAL.
SQ   SEQUENCE   723 AA;  77956 MW;  A1D48BDB CRC32;

  Query Match          27.7%;  Score 321;  DB 1;  Length 723;
  Best Local Similarity  43.8%;  Pred. No. 2.05e-49;
  Matches   42;  Conservative   20;  Mismatches  31;  Indels   3;  Gaps   2;

Db    429 CRCQAGFSGRHCDDNVDDCASSPCANGGTCRDGVNDFSCTCPPGYTGRNCS-APVSR--C 485
          | |  |::| :| :| |||    | ||: | | || ::| |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    486 EHAPCHNGATCHERGHGYVCECARGYGGPNCQFLLP 521
          | : |:||| | ::|   ||:|  |:|||:|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    5
ID   DLL1_RAT       STANDARD;      PRT;   714 AA.
AC   P97677;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DISIBIO G., HEBSHI L., BOULTER J., WEINMASTER G.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78889; G1699046; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   714       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   537       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   538   560       POTENTIAL.
FT   DOMAIN     561   714       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225   253       EGF-LIKE 1.
FT   DOMAIN     256   284       EGF-LIKE 2.
FT   DOMAIN     291   324       EGF-LIKE 3.
FT   DOMAIN     331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369   401       EGF-LIKE 5.
FT   DOMAIN     408   439       EGF-LIKE 6.
FT   DOMAIN     446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484   515       EGF-LIKE 8.
FT   DISULFID   225   236       BY SIMILARITY.
FT   DISULFID   229   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   262   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   303       BY SIMILARITY.
FT   DISULFID   297   313       BY SIMILARITY.
FT   DISULFID   315   324       BY SIMILARITY.
FT   DISULFID   331   342       BY SIMILARITY.
FT   DISULFID   336   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   380       BY SIMILARITY.
FT   DISULFID   374   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   CARBOHYD   476   476       POTENTIAL.
SQ   SEQUENCE   714 AA;  77378 MW;  604B76D1 CRC32;

  Query Match          27.2%;  Score 315;  DB 1;  Length 714;
  Best Local Similarity  42.7%;  Pred. No. 5.03e-48;
  Matches   41;  Conservative   21;  Mismatches  31;  Indels   3;  Gaps   2;

Db    428 CRCQTGFSGRYCEDNVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGRNCS-APVSR--C 484
          | |  |::|  | :| |||    | ||: | | || ::| |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68
```

```
Db    485 EHAPCHNGATCHQRGQRYMCECAQGYGGANCQFLLP 520
          | : |:||| | ::| | :|:|  |:||::|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    6
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   570       FIBROPELLIN C.
FT   DOMAIN      18    55       EGF-LIKE 1.
FT   DOMAIN      62   175       CUB.
FT   DOMAIN     176   212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214   250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252   288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290   326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328   364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366   402       EGF-LIKE 7.
FT   DOMAIN     404   440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442   570       AVIDIN-LIKE.
FT   DISULFID    23    34       BY SIMILARITY.
FT   DISULFID    28    43       BY SIMILARITY.
FT   DISULFID    45    54       BY SIMILARITY.
FT   DISULFID   180   191       BY SIMILARITY.
FT   DISULFID   185   200       BY SIMILARITY.
FT   DISULFID   202   211       BY SIMILARITY.
FT   DISULFID   218   229       BY SIMILARITY.
FT   DISULFID   223   238       BY SIMILARITY.
FT   DISULFID   240   249       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   261   276       BY SIMILARITY.
FT   DISULFID   278   287       BY SIMILARITY.
FT   DISULFID   294   305       BY SIMILARITY.
FT   DISULFID   299   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   CARBOHYD    30    30       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match           26.7%;  Score 310;  DB 1;  Length 570;
  Best Local Similarity  42.4%;  Pred. No. 7.18e-47;
  Matches   39;  Conservative   19;  Mismatches  31;  Indels   3;  Gaps   2;

Db    314 CDCRAGFTGSNCETNINECASSPCLNGGSCLDGVDGYVCQCLPNYTGTHCEIS-L--DAC 370
          |:|  |:|| ||  | ::|    | ||: |:| |::| | |:  |:|  |||:     :|
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    371 ASLPCQNGGVCTNVGGDYVCECLPGYTGINCE 402
            :  ||||: | : |:  ||:||||: | :||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    7
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
```

```
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     44       POTENTIAL.
FT   CHAIN       45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746   1766       POTENTIAL.
FT   DOMAIN    1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58     95       EGF-LIKE 1.
FT   DOMAIN      96    136       EGF-LIKE 2.
FT   DOMAIN     139    176       EGF-LIKE 3.
FT   DOMAIN     177    215       EGF-LIKE 4.
FT   DOMAIN     217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255    291       EGF-LIKE 6.
FT   DOMAIN     293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     409    447       EGF-LIKE 10.
FT   DOMAIN     449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     867    905       EGF-LIKE 22.
FT   DOMAIN     907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 25.
FT   DOMAIN    1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1060   1096       EGF-LIKE 27.
FT   DOMAIN    1098   1134       EGF-LIKE 28.
FT   DOMAIN    1136   1181       EGF-LIKE 29.
FT   DOMAIN    1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1297   1335       EGF-LIKE 33.
FT   DOMAIN    1337   1373       EGF-LIKE 34.
FT   DOMAIN    1375   1412       EGF-LIKE 35.
FT   DOMAIN    1415   1451       EGF-LIKE 36.
FT   DOMAIN    1475   1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1475   1513       LIN/NOTCH 1.
FT   REPEAT    1514   1553       LIN/NOTCH 2.
FT   REPEAT    1554   1593       LIN/NOTCH 3.
FT   DOMAIN    1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN    2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     83       BY SIMILARITY.
FT   DISULFID    85     94       BY SIMILARITY.
FT   DISULFID   100    111       BY SIMILARITY.
FT   DISULFID   105    124       BY SIMILARITY.
FT   DISULFID   126    135       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   181    192       BY SIMILARITY.
FT   DISULFID   186    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    241       BY SIMILARITY.
FT   DISULFID   243    252       BY SIMILARITY.
FT   DISULFID   259    270       BY SIMILARITY.
FT   DISULFID   264    279       BY SIMILARITY.
FT   DISULFID   281    290       BY SIMILARITY.
FT   DISULFID   297    308       BY SIMILARITY.
FT   DISULFID   302    317       BY SIMILARITY.
FT   DISULFID   319    328       BY SIMILARITY.
FT   DISULFID   335    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    396       BY SIMILARITY.
FT   DISULFID   398    407       BY SIMILARITY.
FT   DISULFID   413    424       BY SIMILARITY.
FT   DISULFID   418    435       BY SIMILARITY.
FT   DISULFID   437    446       BY SIMILARITY.
FT   DISULFID   453    465       BY SIMILARITY.
FT   DISULFID   459    474       BY SIMILARITY.
FT   DISULFID   476    485       BY SIMILARITY.
FT   DISULFID   492    503       BY SIMILARITY.
FT   DISULFID   497    512       BY SIMILARITY.
FT   DISULFID   514    523       BY SIMILARITY.
FT   DISULFID   530    541       BY SIMILARITY.
FT   DISULFID   535    550       BY SIMILARITY.
FT   DISULFID   552    561       BY SIMILARITY.
FT   DISULFID   568    579       BY SIMILARITY.
```

```
FT   DISULFID    573    588       BY SIMILARITY.
FT   DISULFID    590    599       BY SIMILARITY.
FT   DISULFID    606    616       BY SIMILARITY.
FT   DISULFID    611    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    654       BY SIMILARITY.
FT   DISULFID    648    663       BY SIMILARITY.
FT   DISULFID    665    674       BY SIMILARITY.
FT   DISULFID    681    692       BY SIMILARITY.
FT   DISULFID    686    701       BY SIMILARITY.
FT   DISULFID    703    712       BY SIMILARITY.
FT   DISULFID    719    730       BY SIMILARITY.
FT   DISULFID    724    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
..
Note: remainder of annotations omitted.

  Query Match          25.9%;  Score 301;  DB 1;  Length 2703;
  Best Local Similarity  40.7%;  Pred. No. 8.47e-45;
  Matches   37;  Conservative   19;  Mismatches  32;  Indels   3;  Gaps   2;

Db   1008 CTCPLGFSGINCQTNDEDCTESSCLNGGSCIDGINGYNCSCLAGYSGANCQYKLN-K--C 1064
           | | ||::| ||  |::|| :  | ||: |:| :|:| | |: |||  |:     :  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db   1065 DSNPCLNGATCHEQNNEYTCHCPSGFTGKQC 1095
           ::  | ||| | :| :   |:| :|| | :|
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPEC 99


RESULT    8
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
     CYPRINIDAE; RASBORINAE; DANIO.
     [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN       21   1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1725   1747       POTENTIAL.
FT   DOMAIN     1748   2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       21     57       EGF-LIKE 1.
FT   DOMAIN       58     98       EGF-LIKE 2.
FT   DOMAIN      101    138       EGF-LIKE 3.
FT   DOMAIN      139    175       EGF-LIKE 4.
FT   DOMAIN      177    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      217    254       EGF-LIKE 6.
FT   DOMAIN      256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      371    409       EGF-LIKE 10.
FT   DOMAIN      411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      827    865       EGF-LIKE 22.
FT   DOMAIN      867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      981   1017       EGF-LIKE 26.
FT   DOMAIN     1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1057   1093       EGF-LIKE 28.
FT   DOMAIN     1095   1141       EGF-LIKE 29.
FT   DOMAIN     1143   1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1181   1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1219   1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1265   1303       EGF-LIKE 33.
FT   DOMAIN     1305   1344       EGF-LIKE 34.
FT   DOMAIN     1346   1382       EGF-LIKE 35.
FT   DOMAIN     1385   1423       EGF-LIKE 36.
FT   DOMAIN     1446   1561       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1486       LIN/NOTCH 1.
FT   REPEAT     1487   1520       LIN/NOTCH 2.
FT   REPEAT     1521   1561       LIN/NOTCH 3.
FT   DOMAIN     1861   2074       6 X ANK MOTIF REPEATS.
FT   REPEAT     1861   1891       ANK MOTIF 1.
FT   REPEAT     1892   1940       ANK MOTIF 1.
FT   REPEAT     1941   1974       ANK MOTIF 1.
FT   REPEAT     1975   2007       ANK MOTIF 1.
FT   REPEAT     2008   2040       ANK MOTIF 1.
FT   REPEAT     2041   2074       ANK MOTIF 1.
```

```
FT   DOMAIN     2265   2276       POLY-GLN (OPA-REPEAT).
FT   DISULFID     25     35       BY SIMILARITY.
FT   DISULFID     29     45       BY SIMILARITY.
FT   DISULFID     47     56       BY SIMILARITY.
FT   DISULFID     62     73       BY SIMILARITY.
FT   DISULFID     67     86       BY SIMILARITY.
FT   DISULFID     88     97       BY SIMILARITY.
FT   DISULFID    105    116       BY SIMILARITY.
FT   DISULFID    110    126       BY SIMILARITY.
FT   DISULFID    128    137       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    163       BY SIMILARITY.
FT   DISULFID    165    174       BY SIMILARITY.
FT   DISULFID    181    194       BY SIMILARITY.
FT   DISULFID    188    203       BY SIMILARITY.
FT   DISULFID    205    214       BY SIMILARITY.
FT   DISULFID    221    232       BY SIMILARITY.
FT   DISULFID    226    242       BY SIMILARITY.
FT   DISULFID    244    253       BY SIMILARITY.
FT   DISULFID    260    271       BY SIMILARITY.
FT   DISULFID    265    280       BY SIMILARITY.
FT   DISULFID    282    291       BY SIMILARITY.
FT   DISULFID    298    311       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    322    331       BY SIMILARITY.
FT   DISULFID    338    349       BY SIMILARITY.
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    375    386       BY SIMILARITY.
FT   DISULFID    380    397       BY SIMILARITY.
FT   DISULFID    399    408       BY SIMILARITY.
FT   DISULFID    415    428       BY SIMILARITY.
FT   DISULFID    422    437       BY SIMILARITY.
FT   DISULFID    439    448       BY SIMILARITY.
FT   DISULFID    455    466       BY SIMILARITY.
FT   DISULFID    460    475       BY SIMILARITY.
FT   DISULFID    477    486       BY SIMILARITY.
FT   DISULFID    493    503       BY SIMILARITY.
FT   DISULFID    498    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    578       BY SIMILARITY.
FT   DISULFID    573    587       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    605    616       BY SIMILARITY.
FT   DISULFID    610    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    653       BY SIMILARITY.
FT   DISULFID    648    662       BY SIMILARITY.
FT   DISULFID    664    673       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   DISULFID    685    700       BY SIMILARITY.
FT   DISULFID    702    711       BY SIMILARITY.
FT   DISULFID    718    728       BY SIMILARITY.
FT   DISULFID    723    737       BY SIMILARITY.
FT   DISULFID    739    748       BY SIMILARITY.
FT   DISULFID    755    766       BY SIMILARITY.
FT   DISULFID    760    775       BY SIMILARITY.
FT   DISULFID    777    786       BY SIMILARITY.
FT   DISULFID    793    804       BY SIMILARITY.
FT   DISULFID    798    813       BY SIMILARITY.
FT   DISULFID    815    824       BY SIMILARITY.
FT   DISULFID    831    842       BY SIMILARITY.
FT   DISULFID    836    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
FT   DISULFID    871    882       BY SIMILARITY.
FT   DISULFID    876    891       BY SIMILARITY.
FT   DISULFID    893    902       BY SIMILARITY.
FT   DISULFID    909    920       BY SIMILARITY.
FT   DISULFID    914    929       BY SIMILARITY.
FT   DISULFID    931    940       BY SIMILARITY.
FT   DISULFID    947    958       BY SIMILARITY.
FT   DISULFID    952    967       BY SIMILARITY.
FT   DISULFID    969    978       BY SIMILARITY.
FT   DISULFID   1023   1034       BY SIMILARITY.
FT   DISULFID   1028   1043       BY SIMILARITY.
FT   DISULFID   1045   1054       BY SIMILARITY.
FT   DISULFID   1061   1072       BY SIMILARITY.
FT   DISULFID   1066   1081       BY SIMILARITY.
FT   DISULFID   1083   1092       BY SIMILARITY.
FT   DISULFID   1099   1120       BY SIMILARITY.
FT   DISULFID   1114   1129       BY SIMILARITY.
FT   DISULFID   1131   1140       BY SIMILARITY.
FT   DISULFID   1147   1158       BY SIMILARITY.
FT   DISULFID   1152   1167       BY SIMILARITY.
FT   DISULFID   1169   1178       BY SIMILARITY.
FT   DISULFID   1185   1196       BY SIMILARITY.
FT   DISULFID   1190   1205       BY SIMILARITY.
FT   DISULFID   1207   1216       BY SIMILARITY.
FT   DISULFID   1223   1242       BY SIMILARITY.
FT   DISULFID   1236   1251       BY SIMILARITY.
FT   DISULFID   1253   1262       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match           25.8%;  Score 299;  DB 1;  Length 2437;
  Best Local Similarity  44.1%;  Pred. No. 2.44e-44;
  Matches   41;  Conservative   14;  Mismatches  34;  Indels   4;  Gaps   3;

Db    474 HCICMPGYEGVFCQINSDDCASQPCLNG-KCIDKINSFHCECPKGFSGSLCQVD-V--DE 529
          :| || || |  |  | |||  : | ||  |:| :||: | |  |:|| ||::
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    530 CASTPCKNGAKCTDGPNKYTCECTPGFSGIHCE 562
          | :| | |||:| |  ::  |:| |||:|  ||
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    9
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643       EXTRACELLULAR.
FT   TRANSMEM  1644   1664       POTENTIAL.
FT   DOMAIN    1665   2318       CYTOPLASMIC.
FT   DOMAIN      39   1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261       PEST.
FT   DOMAIN      39     78       EGF-LIKE 1.
FT   DOMAIN      79    119       EGF-LIKE 2.
FT   DOMAIN     120    157.      EGF-LIKE 3.
FT   DOMAIN     159    196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235       EGF-LIKE 5.
FT   DOMAIN     237    273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313       EGF-LIKE 7.
FT   DOMAIN     315    351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390       EGF-LIKE 9.
FT   DOMAIN     392    430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731       EGF-LIKE 18.
FT   DOMAIN     735    771       EGF-LIKE 19.
FT   DOMAIN     772    809       EGF-LIKE 20.
FT   DOMAIN     811    848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961       EGF-LIKE 24.
FT   DOMAIN     963    999       EGF-LIKE 25.
FT   DOMAIN    1001   1035       EGF-LIKE 26.
FT   DOMAIN    1037   1083       EGF-LIKE 27.
FT   DOMAIN    1085   1121       EGF-LIKE 28.
FT   DOMAIN    1123   1159       EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245       EGF-LIKE 31.
FT   DOMAIN    1247   1288       EGF-LIKE 32.
FT   DOMAIN    1290   1326       EGF-LIKE 33.
FT   DOMAIN    1336   1374       EGF-LIKE 34.
FT   REPEAT    1388   1428       LIN/NOTCH 1.
FT   REPEAT    1429   1467       LIN/NOTCH 2.
FT   REPEAT    1468   1503       LIN/NOTCH 3.
FT   REPEAT    1784   1816       CDC10/SWI6 1.
FT   REPEAT    1817   1865       CDC10/SWI6 2.
FT   REPEAT    1866   1898       CDC10/SWI6 3.
FT   REPEAT    1899   1932       CDC10/SWI6 4.
FT   REPEAT    1933   1965       CDC10/SWI6 5.
FT   REPEAT    1966   1998       CDC10/SWI6 6.
FT   DISULFID    43     55       BY SIMILARITY.
FT   DISULFID    49     66       BY SIMILARITY.
FT   DISULFID    68     77       BY SIMILARITY.
FT   DISULFID    83     94       BY SIMILARITY.
FT   DISULFID    88    107       BY SIMILARITY.
FT   DISULFID   109    118       BY SIMILARITY.
FT   DISULFID   124    135       BY SIMILARITY.
FT   DISULFID   129    145       BY SIMILARITY.
FT   DISULFID   147    156       BY SIMILARITY.
FT   DISULFID   163    175       BY SIMILARITY.
FT   DISULFID   169    184       BY SIMILARITY.
FT   DISULFID   186    195       BY SIMILARITY.
FT   DISULFID   202    213       BY SIMILARITY.
FT   DISULFID   207    223       BY SIMILARITY.
FT   DISULFID   225    234       BY SIMILARITY.
FT   DISULFID   241    252       BY SIMILARITY.
FT   DISULFID   246    261       BY SIMILARITY.
FT   DISULFID   263    272       BY SIMILARITY.
FT   DISULFID   279    292       BY SIMILARITY.
FT   DISULFID   286    301       BY SIMILARITY.
FT   DISULFID   303    312       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   324    339       BY SIMILARITY.
FT   DISULFID   341    350       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    378       BY SIMILARITY.
FT   DISULFID   380    389       BY SIMILARITY.
FT   DISULFID   396    409       BY SIMILARITY.
FT   DISULFID   403    418       BY SIMILARITY.
FT   DISULFID   420    429       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   441    456       BY SIMILARITY.
FT   DISULFID   458    467       BY SIMILARITY.
FT   DISULFID   474    485       BY SIMILARITY.
FT   DISULFID   479    494       BY SIMILARITY.
FT   DISULFID   496    505       BY SIMILARITY.
FT   DISULFID   512    523       BY SIMILARITY.
FT   DISULFID   517    532       BY SIMILARITY.
FT   DISULFID   534    543       BY SIMILARITY.
FT   DISULFID   550    560       BY SIMILARITY.
FT   DISULFID   555    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   587    598       BY SIMILARITY.
FT   DISULFID   592    607       BY SIMILARITY.
FT   DISULFID   609    618       BY SIMILARITY.
FT   DISULFID   625    635       BY SIMILARITY.
FT   DISULFID   630    644       BY SIMILARITY.
FT   DISULFID   646    655       BY SIMILARITY.
FT   DISULFID   662    673       BY SIMILARITY.
FT   DISULFID   667    682       BY SIMILARITY.
FT   DISULFID   684    693       BY SIMILARITY.
FT   DISULFID   700    710       BY SIMILARITY.
FT   DISULFID   705    719       BY SIMILARITY.
FT   DISULFID   721    730       BY SIMILARITY.
FT   DISULFID   739    750       BY SIMILARITY.
FT   DISULFID   744    759       BY SIMILARITY.
FT   DISULFID   761    770       BY SIMILARITY.
FT   DISULFID   776    787       BY SIMILARITY.
FT   DISULFID   781    797       BY SIMILARITY.
FT   DISULFID   799    808       BY SIMILARITY.
FT   DISULFID   815    827       BY SIMILARITY.
FT   DISULFID   821    836       BY SIMILARITY.
FT   DISULFID   838    847       BY SIMILARITY.
FT   DISULFID   854    865       BY SIMILARITY.
FT   DISULFID   859    874       BY SIMILARITY.
FT   DISULFID   876    885       BY SIMILARITY.
FT   DISULFID   892    902       BY SIMILARITY.
FT   DISULFID   897    911       BY SIMILARITY.
FT   DISULFID   913    922       BY SIMILARITY.
FT   DISULFID   929    940       BY SIMILARITY.
FT   DISULFID   934    949       BY SIMILARITY.
FT   DISULFID   951    960       BY SIMILARITY.
FT   DISULFID   967    978       BY SIMILARITY.
FT   DISULFID   972    987       BY SIMILARITY.
FT   DISULFID   989    998       BY SIMILARITY.
FT   DISULFID  1005   1016       BY SIMILARITY.
FT   DISULFID  1010   1023       BY SIMILARITY.
FT   DISULFID  1025   1034       BY SIMILARITY.
FT   DISULFID  1041   1062       BY SIMILARITY.
FT   DISULFID  1056   1071       BY SIMILARITY.
FT   DISULFID  1073   1082       BY SIMILARITY.
FT   DISULFID  1089   1100       BY SIMILARITY.
FT   DISULFID  1094   1109       BY SIMILARITY.
```

```
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          24.8%;  Score 288;  DB 1;  Length 2318;
  Best Local Similarity  41.9%;  Pred. No. 8.01e-42;
  Matches   39;  Conservative   16;  Mismatches  35;  Indels   3;  Gaps   2;

Db    456 CICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMCQLD-V--DEC 512
          | || |:|| |  : |:|    | ||: | | ||:::| | |:|| :|::       |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    513 ASTPCRNGAKCVDQPDGYECRCAEGFEGTLCER 545
           :| |:|||:||||     |:|  || |  ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEK 101


RESULT   10
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1     90
FT   CHAIN       91   2139       CRUMBS PROTEIN.
FT   DOMAIN      91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085   2111       POTENTIAL.
FT   DOMAIN    2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267    303       EGF-LIKE 1.
FT   DOMAIN     306    343       EGF-LIKE 2.
FT   DOMAIN     348    386       EGF-LIKE 3.
FT   DOMAIN     388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427    463       EGF-LIKE 5.
FT   DOMAIN     464    500       EGF-LIKE 6.
FT   DOMAIN     501    532       EGF-LIKE 7.
FT   DOMAIN     545    581       EGF-LIKE 8.
FT   DOMAIN     582    611       EGF-LIKE 9.
FT   DOMAIN     609    646       EGF-LIKE 10.
FT   DOMAIN     648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840    902       EGF-LIKE 16.
FT   DOMAIN     904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980   1021       EGF-LIKE 19.
FT   DOMAIN    1207   1243       EGF-LIKE 20.
FT   DOMAIN    1481   1517       EGF-LIKE 21.
FT   DOMAIN    1759   1795       EGF-LIKE 22.
FT   DOMAIN    1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874   1915       EGF-LIKE 25.
FT   DOMAIN    1915   1951       EGF-LIKE 26.
FT   DOMAIN    1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030   2070       EGF-LIKE 29.
FT   DISULFID   271    282       BY SIMILARITY.
FT   DISULFID   276    291       BY SIMILARITY.
FT   DISULFID   293    302       BY SIMILARITY.
FT   DISULFID   310    321       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   333    342       BY SIMILARITY.
FT   DISULFID   352    363       BY SIMILARITY.
FT   DISULFID   357    374       BY SIMILARITY.
FT   DISULFID   376    385       BY SIMILARITY.
FT   DISULFID   392    403       BY SIMILARITY.
FT   DISULFID   397    412       BY SIMILARITY.
FT   DISULFID   414    424       BY SIMILARITY.
FT   DISULFID   431    442       BY SIMILARITY.
FT   DISULFID   436    451       BY SIMILARITY.
FT   DISULFID   453    462       BY SIMILARITY.
FT   DISULFID   468    479       BY SIMILARITY.
FT   DISULFID   473    488       BY SIMILARITY.
FT   DISULFID   490    499       BY SIMILARITY.
```

```
FT   DISULFID   505   515       BY SIMILARITY.
FT   DISULFID   509   520       BY SIMILARITY.
FT   DISULFID   522   531       BY SIMILARITY.
FT   DISULFID   549   562       BY SIMILARITY.
FT   DISULFID   556   569       BY SIMILARITY.
FT   DISULFID   571   580       BY SIMILARITY.
FT   DISULFID   586   597       BY SIMILARITY.
FT   DISULFID   591   602       BY SIMILARITY.
FT   DISULFID   604   610       BY SIMILARITY.
FT   DISULFID   613   624       BY SIMILARITY.
FT   DISULFID   618   634       BY SIMILARITY.
FT   DISULFID   636   645       BY SIMILARITY.
FT   DISULFID   652   664       BY SIMILARITY.
FT   DISULFID   659   673       BY SIMILARITY.
FT   DISULFID   675   684       BY SIMILARITY.
FT   DISULFID   691   702       BY SIMILARITY.
FT   DISULFID   696   711       BY SIMILARITY.
FT   DISULFID   713   722       BY SIMILARITY.
FT   DISULFID   729   740       BY SIMILARITY.
FT   DISULFID   734   749       BY SIMILARITY.
FT   DISULFID   751   760       BY SIMILARITY.
FT   DISULFID   767   778       BY SIMILARITY.
FT   DISULFID   772   787       BY SIMILARITY.
FT   DISULFID   789   799       BY SIMILARITY.
FT   DISULFID   806   817       BY SIMILARITY.
FT   DISULFID   811   826       BY SIMILARITY.
FT   DISULFID   828   837       BY SIMILARITY.
FT   DISULFID   844   855       BY SIMILARITY.
FT   DISULFID   849   890       BY SIMILARITY.
FT   DISULFID   892   901       BY SIMILARITY.
FT   DISULFID   908   919       BY SIMILARITY.
FT   DISULFID   913   928       BY SIMILARITY.
FT   DISULFID   930   939       BY SIMILARITY.
FT   DISULFID   946   957       BY SIMILARITY.
FT   DISULFID   952   966       BY SIMILARITY.
FT   DISULFID   968   977       BY SIMILARITY.
FT   DISULFID   984   995       BY SIMILARITY.
FT   DISULFID   989  1009       BY SIMILARITY.
FT   DISULFID  1011  1020       BY SIMILARITY.
FT   DISULFID  1211  1222       BY SIMILARITY.
FT   DISULFID  1216  1231       BY SIMILARITY.
FT   DISULFID  1233  1242       BY SIMILARITY.
FT   DISULFID  1485  1496       BY SIMILARITY.
FT   DISULFID  1490  1505       BY SIMILARITY.
FT   DISULFID  1507  1516       BY SIMILARITY.
FT   DISULFID  1763  1774       BY SIMILARITY.
FT   DISULFID  1768  1783       BY SIMILARITY.
FT   DISULFID  1785  1794       BY SIMILARITY.
FT   DISULFID  1801  1812       BY SIMILARITY.
FT   DISULFID  1806  1821       BY SIMILARITY.
FT   DISULFID  1823  1832       BY SIMILARITY.
FT   DISULFID  1839  1850       BY SIMILARITY.
FT   DISULFID  1844  1859       BY SIMILARITY.
FT   DISULFID  1861  1870       BY SIMILARITY.
FT   DISULFID  1878  1889       BY SIMILARITY.
FT   DISULFID  1883  1903       BY SIMILARITY.
FT   DISULFID  1905  1914       BY SIMILARITY.
FT   DISULFID  1919  1930       BY SIMILARITY.
FT   DISULFID  1924  1939       BY SIMILARITY.
FT   DISULFID  1941  1950       BY SIMILARITY.
FT   DISULFID  1957  1968       BY SIMILARITY.
FT   DISULFID  1962  1977       BY SIMILARITY.
FT   DISULFID  1979  1988       BY SIMILARITY.
FT   DISULFID  1995  2008       BY SIMILARITY.
FT   DISULFID  2002  2017       BY SIMILARITY.
FT   DISULFID  2019  2028       BY SIMILARITY.
FT   CARBOHYD    37    37       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   198   198       POTENTIAL.
FT   CARBOHYD   238   238       POTENTIAL.
FT   CARBOHYD   239   239       POTENTIAL.
FT   CARBOHYD   336   336       POTENTIAL.
FT   CARBOHYD   400   400       POTENTIAL.
FT   CARBOHYD   550   550       POTENTIAL.
FT   CARBOHYD   565   565       POTENTIAL.
FT   CARBOHYD   736   736       POTENTIAL.
FT   CARBOHYD   746   746       POTENTIAL.
FT   CARBOHYD   860   860       POTENTIAL.
FT   CARBOHYD   884   884       POTENTIAL.
FT   CARBOHYD   976   976       POTENTIAL.
FT   CARBOHYD  1102  1102       POTENTIAL.
FT   CARBOHYD  1114  1114       POTENTIAL.
FT   CARBOHYD  1138  1138       POTENTIAL.
FT   CARBOHYD  1192  1192       POTENTIAL.
FT   CARBOHYD  1245  1245       POTENTIAL.
FT   CARBOHYD  1255  1255       POTENTIAL.
FT   CARBOHYD  1354  1354       POTENTIAL.
FT   CARBOHYD  1363  1363       POTENTIAL.
FT   CARBOHYD  1441  1441       POTENTIAL.
FT   CARBOHYD  1454  1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match           24.3%;  Score 282;  DB 1;  Length 2139;
  Best Local Similarity  39.6%;  Pred. No. 1.86e-40;
  Matches   38;  Conservative   19;  Mismatches  34;  Indels   5;  Gaps   5;

Db   1821 CQCQPGFEGQHCEQNIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTCEN 1880
            |:|  |: |::| :| |:| |: |:||: | | : || | | | | |  |::    :
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEI-PP-A-PR 65

Db   1881 EPCRNGSTCQNGFN-ASTGNNFTCTCVPGFEGPLCD 1915
             :| :|: |||| |    |:   | |:||| || |:
Qy     66 SSC-EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   11
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
```

```
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1   19       POTENTIAL.
FT   CHAIN       20 2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20 1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729 1750       POTENTIAL.
FT   DOMAIN    1751 2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20   57       EGF-LIKE 1.
FT   DOMAIN      58   99       EGF-LIKE 2.
FT   DOMAIN     102  140       EGF-LIKE 3.
FT   DOMAIN     141  177       EGF-LIKE 4.
FT   DOMAIN     179  215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217  254       EGF-LIKE 6.
FT   DOMAIN     256  292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294  332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334  370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371  409       EGF-LIKE 10.
FT   DOMAIN     411  449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451  487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489  525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527  563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565  600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602  638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     640  675       EGF-LIKE 17.
FT   DOMAIN     677  713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715  750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     752  788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     790  826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     828  866       EGF-LIKE 22.
FT   DOMAIN     868  904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     906  942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     944  980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     982 1018       EGF-LIKE 26.
FT   DOMAIN    1020 1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1058 1094       EGF-LIKE 28.
FT   DOMAIN    1096 1142       EGF-LIKE 29.
FT   DOMAIN    1144 1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1182 1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1220 1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1266 1304       EGF-LIKE 33.
FT   DOMAIN    1306 1346       EGF-LIKE 34.
FT   DOMAIN    1347 1383       EGF-LIKE 35.
FT   DOMAIN    1386 1424       EGF-LIKE 36.
FT   DOMAIN    1441 1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1441 1478       LIN/NOTCH 1.
FT   REPEAT    1479 1520       LIN/NOTCH 2.
FT   REPEAT    1521 1560       LIN/NOTCH 3.
FT   DOMAIN    1871 2083       6 X ANK MOTIF REPEATS.
FT   DISULFID    22   35       BY SIMILARITY.
FT   DISULFID    29   45       BY SIMILARITY.
FT   DISULFID    47   56       BY SIMILARITY.
FT   DISULFID    62   74       BY SIMILARITY.
FT   DISULFID    68   87       BY SIMILARITY.
FT   DISULFID    89   98       BY SIMILARITY.
FT   DISULFID   106  117       BY SIMILARITY.
FT   DISULFID   111  128       BY SIMILARITY.
FT   DISULFID   130  139       BY SIMILARITY.
FT   DISULFID   145  156       BY SIMILARITY.
FT   DISULFID   150  165       BY SIMILARITY.
FT   DISULFID   167  176       BY SIMILARITY.
FT   DISULFID   183  194       BY SIMILARITY.
FT   DISULFID   188  203       BY SIMILARITY.
FT   DISULFID   205  214       BY SIMILARITY.
FT   DISULFID   221  232       BY SIMILARITY.
FT   DISULFID   226  242       BY SIMILARITY.
FT   DISULFID   244  253       BY SIMILARITY.
FT   DISULFID   260  271       BY SIMILARITY.
FT   DISULFID   265  280       BY SIMILARITY.
FT   DISULFID   282  291       BY SIMILARITY.
FT   DISULFID   298  311       BY SIMILARITY.
FT   DISULFID   305  320       BY SIMILARITY.
FT   DISULFID   322  331       BY SIMILARITY.
FT   DISULFID   338  349       BY SIMILARITY.
FT   DISULFID   343  358       BY SIMILARITY.
FT   DISULFID   360  369       BY SIMILARITY.
FT   DISULFID   375  386       BY SIMILARITY.
FT   DISULFID   380  397       BY SIMILARITY.
FT   DISULFID   399  408       BY SIMILARITY.
FT   DISULFID   415  428       BY SIMILARITY.
FT   DISULFID   422  437       BY SIMILARITY.
FT   DISULFID   439  448       BY SIMILARITY.
FT   DISULFID   455  466       BY SIMILARITY.
FT   DISULFID   460  475       BY SIMILARITY.
FT   DISULFID   477  486       BY SIMILARITY.
FT   DISULFID   493  504       BY SIMILARITY.
FT   DISULFID   498  513       BY SIMILARITY.
FT   DISULFID   515  524       BY SIMILARITY.
FT   DISULFID   531  542       BY SIMILARITY.
FT   DISULFID   536  551       BY SIMILARITY.
FT   DISULFID   553  562       BY SIMILARITY.
FT   DISULFID   569  579       BY SIMILARITY.
FT   DISULFID   574  588       BY SIMILARITY.
FT   DISULFID   590  599       BY SIMILARITY.
FT   DISULFID   606  617       BY SIMILARITY.
FT   DISULFID   611  626       BY SIMILARITY.
FT   DISULFID   628  637       BY SIMILARITY.
FT   DISULFID   644  654       BY SIMILARITY.
FT   DISULFID   649  663       BY SIMILARITY.
FT   DISULFID   665  674       BY SIMILARITY.
FT   DISULFID   681  692       BY SIMILARITY.
FT   DISULFID   686  701       BY SIMILARITY.
FT   DISULFID   703  712       BY SIMILARITY.
FT   DISULFID   719  729       BY SIMILARITY.
FT   DISULFID   724  738       BY SIMILARITY.
FT   DISULFID   740  749       BY SIMILARITY.
FT   DISULFID   756  767       BY SIMILARITY.
FT   DISULFID   761  776       BY SIMILARITY.
FT   DISULFID   778  787       BY SIMILARITY.
FT   DISULFID   794  805       BY SIMILARITY.
FT   DISULFID   799  814       BY SIMILARITY.
FT   DISULFID   816  825       BY SIMILARITY.
FT   DISULFID   832  843       BY SIMILARITY.
FT   DISULFID   837  854       BY SIMILARITY.
FT   DISULFID   856  865       BY SIMILARITY.
FT   DISULFID   872  883       BY SIMILARITY.
FT   DISULFID   877  892       BY SIMILARITY.
FT   DISULFID   894  903       BY SIMILARITY.
FT   DISULFID   910  921       BY SIMILARITY.
FT   DISULFID   915  930       BY SIMILARITY.
FT   DISULFID   932  941       BY SIMILARITY.
FT   DISULFID   986  997       BY SIMILARITY.
FT   DISULFID   991 1006       BY SIMILARITY.
FT   DISULFID  1008 1017       BY SIMILARITY.
FT   DISULFID  1024 1035       BY SIMILARITY.
FT   DISULFID  1029 1044       BY SIMILARITY.
FT   DISULFID  1046 1055       BY SIMILARITY.
FT   DISULFID  1062 1073       BY SIMILARITY.
FT   DISULFID  1067 1082       BY SIMILARITY.
FT   DISULFID  1084 1093       BY SIMILARITY.
FT   DISULFID  1100 1121       BY SIMILARITY.
FT   DISULFID  1115 1130       BY SIMILARITY.
FT   DISULFID  1132 1141       BY SIMILARITY.
FT   DISULFID  1148 1159       BY SIMILARITY.
FT   DISULFID  1153 1168       BY SIMILARITY.
FT   DISULFID  1170 1179       BY SIMILARITY.
FT   DISULFID  1186 1197       BY SIMILARITY.
FT   DISULFID  1191 1206       BY SIMILARITY.
FT   DISULFID  1208 1217       BY SIMILARITY.
FT   DISULFID  1224 1243       BY SIMILARITY.
FT   DISULFID  1237 1252       BY SIMILARITY.
```

```
FT   DISULFID   1254   1263       BY SIMILARITY.
FT   DISULFID   1270   1283       BY SIMILARITY.
FT   DISULFID   1275   1292       BY SIMILARITY.
FT   DISULFID   1294   1303       BY SIMILARITY.
FT   DISULFID   1310   1321       BY SIMILARITY.
FT   DISULFID   1315   1333       BY SIMILARITY.
FT   DISULFID   1335   1344       BY SIMILARITY.
FT   DISULFID   1351   1362       BY SIMILARITY.
FT   DISULFID   1356   1371       BY SIMILARITY.
FT   DISULFID   1373   1382       BY SIMILARITY.
FT   DISULFID   1390   1401       BY SIMILARITY.
FT   DISULFID   1395   1412       BY SIMILARITY.
FT   DISULFID   1414   1423       BY SIMILARITY.
FT   CARBOHYD    462    462       POTENTIAL.
FT   CARBOHYD    887    887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match           24.3%;  Score 282;  DB 1;  Length 2524;
  Best Local Similarity  35.0%;  Pred. No. 1.86e-40;
  Matches    35;  Conservative   25;  Mismatches  36;  Indels   4;  Gaps   2;

Db    158 PFEIQYICKCPPGFHGATCKQDINECSQNPCKNGGQCINEFGSYRCTCQNRFTGRNCDEP 217
          |: ::  | |  |: |  | :: ::| :: | ||:||::|  || | | : ::|: |: |
Qy      2 PLPVHHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIP 61

Db    218 YVP---CNPSPCLNGGTCRQTDDTSYDCTCLPGFSGQNCE 254
          |   |: : | ||: | :    :  | |||||:| :||
Qy     62 PAPRSSCEGTECQNGANCVDQGSRPV-CQCLPGFGGPECE 100


RESULT  12
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
FT   DOMAIN     869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907    943       EGF-LIKE 24.
FT   DOMAIN     945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983   1019       EGF-LIKE 26.
FT   DOMAIN    1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059   1095       EGF-LIKE 28.
FT   DOMAIN    1097   1143       EGF-LIKE 29.
FT   DOMAIN    1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267   1305       EGF-LIKE 33.
FT   DOMAIN    1307   1346       EGF-LIKE 34.
FT   DOMAIN    1348   1384       EGF-LIKE 35.
FT   DOMAIN    1387   1426       EGF-LIKE 36.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1865   2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2076       ANK MOTIF 6.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
```

```
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   261   272       BY SIMILARITY.
FT   DISULFID   266   281       BY SIMILARITY.
FT   DISULFID   283   292       BY SIMILARITY.
FT   DISULFID   299   312       BY SIMILARITY.
FT   DISULFID   306   321       BY SIMILARITY.
FT   DISULFID   323   332       BY SIMILARITY.
FT   DISULFID   339   350       BY SIMILARITY.
FT   DISULFID   344   359       BY SIMILARITY.
FT   DISULFID   361   370       BY SIMILARITY.
FT   DISULFID   376   387       BY SIMILARITY.
FT   DISULFID   381   398       BY SIMILARITY.
FT   DISULFID   400   409       BY SIMILARITY.
FT   DISULFID   416   429       BY SIMILARITY.
FT   DISULFID   423   438       BY SIMILARITY.
FT   DISULFID   440   449       BY SIMILARITY.
FT   DISULFID   456   467       BY SIMILARITY.
FT   DISULFID   461   476       BY SIMILARITY.
FT   DISULFID   478   487       BY SIMILARITY.
FT   DISULFID   494   505       BY SIMILARITY.
FT   DISULFID   499   514       BY SIMILARITY.
FT   DISULFID   516   525       BY SIMILARITY.
FT   DISULFID   532   543       BY SIMILARITY.
FT   DISULFID   537   552       BY SIMILARITY.
FT   DISULFID   554   563       BY SIMILARITY.
FT   DISULFID   570   580       BY SIMILARITY.
FT   DISULFID   575   589       BY SIMILARITY.
FT   DISULFID   591   600       BY SIMILARITY.
FT   DISULFID   607   618       BY SIMILARITY.
FT   DISULFID   612   627       BY SIMILARITY.
FT   DISULFID   629   638       BY SIMILARITY.
FT   DISULFID   645   655       BY SIMILARITY.
FT   DISULFID   650   664       BY SIMILARITY.
FT   DISULFID   666   675       BY SIMILARITY.
FT   DISULFID   682   693       BY SIMILARITY.
FT   DISULFID   687   702       BY SIMILARITY.
FT   DISULFID   704   713       BY SIMILARITY.
FT   DISULFID   720   730       BY SIMILARITY.
FT   DISULFID   725   739       BY SIMILARITY.
FT   DISULFID   741   750       BY SIMILARITY.
FT   DISULFID   757   768       BY SIMILARITY.
FT   DISULFID   762   777       BY SIMILARITY.
FT   DISULFID   779   788       BY SIMILARITY.
FT   DISULFID   795   806       BY SIMILARITY.
FT   DISULFID   800   815       BY SIMILARITY.
FT   DISULFID   817   826       BY SIMILARITY.
FT   DISULFID   833   844       BY SIMILARITY.
FT   DISULFID   838   855       BY SIMILARITY.
FT   DISULFID   857   866       BY SIMILARITY.
FT   DISULFID   873   884       BY SIMILARITY.
FT   DISULFID   878   893       BY SIMILARITY.
FT   DISULFID   895   904       BY SIMILARITY.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   931       BY SIMILARITY.
FT   DISULFID   933   942       BY SIMILARITY.
FT   DISULFID   987   998       BY SIMILARITY.
FT   DISULFID   992  1007       BY SIMILARITY.
FT   DISULFID  1009  1018       BY SIMILARITY.
FT   DISULFID  1025  1036       BY SIMILARITY.
FT   DISULFID  1030  1045       BY SIMILARITY.
FT   DISULFID  1047  1056       BY SIMILARITY.
FT   DISULFID  1063  1074       BY SIMILARITY.
FT   DISULFID  1068  1083       BY SIMILARITY.
FT   DISULFID  1085  1094       BY SIMILARITY.
FT   DISULFID  1101  1122       BY SIMILARITY.
FT   DISULFID  1116  1131       BY SIMILARITY.
FT   DISULFID  1133  1142       BY SIMILARITY.
FT   DISULFID  1149  1160       BY SIMILARITY.
FT   DISULFID  1154  1169       BY SIMILARITY.
FT   DISULFID  1171  1180       BY SIMILARITY.
FT   DISULFID  1187  1198       BY SIMILARITY.
FT   DISULFID  1192  1207       BY SIMILARITY.
FT   DISULFID  1209  1218       BY SIMILARITY.
FT   DISULFID  1225  1244       BY SIMILARITY.
FT   DISULFID  1238  1253       BY SIMILARITY.
FT   DISULFID  1255  1264       BY SIMILARITY.
FT   DISULFID  1271  1284       BY SIMILARITY.
FT   DISULFID  1276  1293       BY SIMILARITY.
FT   DISULFID  1295  1304       BY SIMILARITY.
FT   DISULFID  1311  1322       BY SIMILARITY.
FT   DISULFID  1316  1334       BY SIMILARITY.
FT   DISULFID  1336  1345       BY SIMILARITY.
FT   DISULFID  1352  1363       BY SIMILARITY.
FT   DISULFID  1357  1372       BY SIMILARITY.
FT   DISULFID  1374  1383       BY SIMILARITY.
FT   DISULFID  1391  1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          24.3%;  Score 282;  DB 1;  Length 2531;
  Best Local Similarity  37.0%;  Pred. No. 1.86e-40;
  Matches   34;  Conservative   21;  Mismatches  34;  Indels   3;  Gaps   2;

Db    438 CQCLQGYTGPRCEIDVNECISNPCQNDATCLDQIGEFQCICMPGYEGVYCEIN-T--DEC 494
          |:|: |||  |  : ::|  : ||| | |:|::  : |:|: || |  |||  :    |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    495 ASSPCLHNGRCVDKINEFLCQCPKGFSGHLCQ 526
           :: | : : |||  :  :|||  ||:|  |:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT  13
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2075       ANK MOTIF 6.
FT   CARBOHYD   888    888       POTENTIAL.
FT   CARBOHYD   959    959       POTENTIAL.
FT   CARBOHYD  1179   1179       POTENTIAL.
FT   CARBOHYD  1241   1241       POTENTIAL.
FT   CARBOHYD  1489   1489       POTENTIAL.
FT   CARBOHYD  1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match           24.1%;  Score 279;  DB 1;  Length 2531;
  Best Local Similarity  41.3%;  Pred. No. 8.95e-40;
  Matches   38;  Conservative   17;  Mismatches  34;  Indels   3;  Gaps   2;

Db    931 CDCLPGFQGAFCEEDINECASNPCQNGANCTDCVDSYTCTCPVGFNGIHCE-NNTP--DC 987
          |:|: |: |  | |: ::|  : ||||| | | |:||:| |  |::|  ||   :|   |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    988 TESSCFNGGTCVDGINSFTCLCPPGFTGSYCQ 1019
            : | ||: |||  :   | | ||| |: |:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   14
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19  >2444       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737   1757       POTENTIAL.
FT   DOMAIN    1758  >2444       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20.
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    868       EGF-LIKE 22.
FT   DOMAIN     870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     908    944       EGF-LIKE 24.
FT   DOMAIN     946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 26.
FT   DOMAIN    1022   1058       EGF-LIKE 27.
FT   DOMAIN    1060   1096       EGF-LIKE 28.
FT   DOMAIN    1098   1144       EGF-LIKE 29.
FT   DOMAIN    1146   1182       EGF-LIKE 30.
FT   DOMAIN    1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1268   1306       EGF-LIKE 33.
FT   DOMAIN    1308   1347       EGF-LIKE 34.
FT   DOMAIN    1349   1385       EGF-LIKE 35.
FT   DOMAIN    1388   1427       EGF-LIKE 36.
FT   DOMAIN    1446   1563       3 X LIN/NOTCH REPEATS.
```

```
FT   REPEAT    1446   1481       LIN/NOTCH 1.
FT   REPEAT    1482   1523       LIN/NOTCH 2.
FT   REPEAT    1524   1563       LIN/NOTCH 3.
FT   DOMAIN    1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT    1876   1921       ANK MOTIF 1.
FT   REPEAT    1923   1954       ANK MOTIF 2.
FT   REPEAT    1956   1987       ANK MOTIF 3.
FT   REPEAT    1990   2021       ANK MOTIF 4.
FT   REPEAT    2023   2054       ANK MOTIF 5.
FT   REPEAT    2056   2087       ANK MOTIF 6.
FT   DOMAIN    1576   1579       POLY-VAL.
FT   DOMAIN    1662   1665       POLY-ARG.
FT   DOMAIN    1729   1732       POLY-PRO.
FT   DOMAIN    1741   1744       POLY-ALA.
FT   DOMAIN    1902   1905       POLY-GLU.
FT   DOMAIN    2260   2263       POLY-GLY.
FT   DOMAIN    2404   2407       POLY-GLN.
FT   DOMAIN    2411   2418       POLY-PRO.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
FT   DISULFID   339    350       BY SIMILARITY.
FT   DISULFID   344    359       BY SIMILARITY.
FT   DISULFID   361    370       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    398       BY SIMILARITY.
FT   DISULFID   400    409       BY SIMILARITY.
FT   DISULFID   416    429       BY SIMILARITY.
FT   DISULFID   423    438       BY SIMILARITY.
FT   DISULFID   440    449       BY SIMILARITY.
FT   DISULFID   456    467       BY SIMILARITY.
FT   DISULFID   461    476       BY SIMILARITY.
FT   DISULFID   478    487       BY SIMILARITY.
FT   DISULFID   494    505       BY SIMILARITY.
FT   DISULFID   499    514       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   532    543       BY SIMILARITY.
FT   DISULFID   537    552       BY SIMILARITY.
FT   DISULFID   554    563       BY SIMILARITY.
FT   DISULFID   570    580       BY SIMILARITY.
FT   DISULFID   575    589       BY SIMILARITY.
FT   DISULFID   591    600       BY SIMILARITY.
FT   DISULFID   607    618       BY SIMILARITY.
FT   DISULFID   612    627       BY SIMILARITY.
FT   DISULFID   629    638       BY SIMILARITY.
FT   DISULFID   645    655       BY SIMILARITY.
FT   DISULFID   650    664       BY SIMILARITY.
FT   DISULFID   666    675       BY SIMILARITY.
FT   DISULFID   682    693       BY SIMILARITY.
FT   DISULFID   687    702       BY SIMILARITY.
FT   DISULFID   704    713       BY SIMILARITY.
FT   DISULFID   720    730       BY SIMILARITY.
FT   DISULFID   725    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    855       BY SIMILARITY.
FT   DISULFID   857    867       BY SIMILARITY.
FT   DISULFID   874    885       BY SIMILARITY.
FT   DISULFID   879    894       BY SIMILARITY.
FT   DISULFID   896    905       BY SIMILARITY.
FT   DISULFID   912    923       BY SIMILARITY.
FT   DISULFID   917    932       BY SIMILARITY.
FT   DISULFID   934    943       BY SIMILARITY.
FT   DISULFID   988    999       BY SIMILARITY.
FT   DISULFID   993   1008       BY SIMILARITY.
FT   DISULFID  1010   1019       BY SIMILARITY.
FT   DISULFID  1026   1037       BY SIMILARITY.
FT   DISULFID  1031   1046       BY SIMILARITY.
FT   DISULFID  1048   1057       BY SIMILARITY.
FT   DISULFID  1064   1075       BY SIMILARITY.
FT   DISULFID  1069   1084       BY SIMILARITY.
FT   DISULFID  1086   1095       BY SIMILARITY.
FT   DISULFID  1102   1123       BY SIMILARITY.
FT   DISULFID  1117   1132       BY SIMILARITY.
FT   DISULFID  1134   1143       BY SIMILARITY.
FT   DISULFID  1150   1161       BY SIMILARITY.
FT   DISULFID  1155   1170       BY SIMILARITY.
FT   DISULFID  1172   1181       BY SIMILARITY.
FT   DISULFID  1188   1199       BY SIMILARITY.
FT   DISULFID  1193   1208       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          24.0%;  Score 278;  DB 1;  Length 2444;
  Best Local Similarity  39.6%;  Pred. No. 1.51e-39;
  Matches   38;  Conservative   19;  Mismatches  36;  Indels   3;  Gaps   2;

Db    932 CDCLPGFRGTFCEEDINECASDPCRNGANCTDCVDSYTCTCPAGFSGIHCE-NNTP--DC 988
          |:|: |: | | |: ::|    |:||| | | |:||:| | |:|| || :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    989 TESSCFNGGTCVDGINSFTCLCPPGFTGSYCQHVVN 1024
            : | ||: ||| :  | | ||| |: |: :::
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT  15
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. VIROL. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
```

RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG, AND TESTIS;
RX MEDLINE; 96281668.
RA UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene.";
RL DEVELOPMENT 122:2251-2259(1996).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M80456; G1714084; -.
DR EMBL; U43691; G1401160; -.
DR PIR; A38072; TVMVT3.
DR MGD; MGI:107471; NOTCH4.
DR PROSITE; PS00010; ASX_HYDROXYL; 11.
DR PROSITE; PS00022; EGF_1; 28.
DR PROSITE; PS01186; EGF_2; 21.
DR PROSITE; PS01187; EGF_CA; 9.
DR PFAM; PF00008; EGF; 26.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 2.
DR HSSP; P00740; 1IXA.
KW DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.

FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	1964	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT	DOMAIN	21	1443	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1444	1464	POTENTIAL.
FT	DOMAIN	1465	1964	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	21	60	EGF-LIKE 1.
FT	DOMAIN	61	112	EGF-LIKE 2.
FT	DOMAIN	115	152	EGF-LIKE 3.
FT	DOMAIN	153	189	EGF-LIKE 4.
FT	DOMAIN	191	229	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	231	271	EGF-LIKE 6.
FT	DOMAIN	273	309	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	311	350	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	388	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	389	427	EGF-LIKE 10.
FT	DOMAIN	429	470	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	472	508	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	510	546	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	548	584	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	586	622	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	622	656	EGF-LIKE 16.
FT	DOMAIN	658	686	EGF-LIKE 17.
FT	DOMAIN	688	724	EGF-LIKE 18.
FT	DOMAIN	726	762	EGF-LIKE 19.
FT	DOMAIN	764	800	EGF-LIKE 20.
FT	DOMAIN	803	839	EGF-LIKE 21.
FT	DOMAIN	841	877	EGF-LIKE 22.
FT	DOMAIN	878	924	EGF-LIKE 23.
FT	DOMAIN	926	962	EGF-LIKE 24.
FT	DOMAIN	964	1000	EGF-LIKE 25.
FT	DOMAIN	1002	1040	EGF-LIKE 26.
FT	DOMAIN	1042	1081	EGF-LIKE 27.
FT	DOMAIN	1083	1122	EGF-LIKE 28.
FT	DOMAIN	1126	1167	EGF-LIKE 29.
FT	DOMAIN	1168	1282	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1168	1208	LIN/NOTCH 1.
FT	REPEAT	1209	1242	LIN/NOTCH 2.
FT	REPEAT	1243	1282	LIN/NOTCH 3.
FT	DOMAIN	1572	1785	6 X ANK MOTIF REPEATS.
FT	REPEAT	1572	1603	ANK MOTIF 1.
FT	REPEAT	1622	1653	ANK MOTIF 2.
FT	REPEAT	1654	1685	ANK MOTIF 3.
FT	REPEAT	1688	1719	ANK MOTIF 4.
FT	REPEAT	1721	1752	ANK MOTIF 5.
FT	REPEAT	1754	1785	ANK MOTIF 6.
FT	DISULFID	25	38	BY SIMILARITY.
FT	DISULFID	32	48	BY SIMILARITY.
FT	DISULFID	50	59	BY SIMILARITY.
FT	DISULFID	65	77	BY SIMILARITY.
FT	DISULFID	71	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	130	BY SIMILARITY.
FT	DISULFID	124	140	BY SIMILARITY.
FT	DISULFID	142	151	BY SIMILARITY.
FT	DISULFID	157	168	BY SIMILARITY.
FT	DISULFID	162	177	BY SIMILARITY.
FT	DISULFID	179	188	BY SIMILARITY.
FT	DISULFID	195	208	BY SIMILARITY.
FT	DISULFID	202	217	BY SIMILARITY.
FT	DISULFID	219	228	BY SIMILARITY.
FT	DISULFID	235	246	BY SIMILARITY.
FT	DISULFID	240	259	BY SIMILARITY.
FT	DISULFID	261	270	BY SIMILARITY.
FT	DISULFID	235	246	BY SIMILARITY.
FT	DISULFID	240	259	BY SIMILARITY.
FT	DISULFID	261	270	BY SIMILARITY.
FT	DISULFID	277	288	BY SIMILARITY.
FT	DISULFID	282	297	BY SIMILARITY.
FT	DISULFID	299	308	BY SIMILARITY.
FT	DISULFID	315	329	BY SIMILARITY.
FT	DISULFID	323	338	BY SIMILARITY.
FT	DISULFID	340	349	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	376	BY SIMILARITY.
FT	DISULFID	378	387	BY SIMILARITY.
FT	DISULFID	393	404	BY SIMILARITY.
FT	DISULFID	398	415	BY SIMILARITY.
FT	DISULFID	417	426	BY SIMILARITY.
FT	DISULFID	433	449	BY SIMILARITY.
FT	DISULFID	443	458	BY SIMILARITY.
FT	DISULFID	460	469	BY SIMILARITY.
FT	DISULFID	476	487	BY SIMILARITY.
FT	DISULFID	481	496	BY SIMILARITY.
FT	DISULFID	498	507	BY SIMILARITY.
FT	DISULFID	514	525	BY SIMILARITY.
FT	DISULFID	519	534	BY SIMILARITY.
FT	DISULFID	536	545	BY SIMILARITY.
FT	DISULFID	552	563	BY SIMILARITY.
FT	DISULFID	557	572	BY SIMILARITY.
FT	DISULFID	574	583	BY SIMILARITY.
FT	DISULFID	590	601	BY SIMILARITY.
FT	DISULFID	595	610	BY SIMILARITY.
FT	DISULFID	612	621	BY SIMILARITY.
FT	DISULFID	626	637	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	648	655	BY SIMILARITY.
FT	DISULFID	662	669	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	692	703	BY SIMILARITY.
FT	DISULFID	697	712	BY SIMILARITY.
FT	DISULFID	714	723	BY SIMILARITY.
FT	DISULFID	730	741	BY SIMILARITY.
FT	DISULFID	735	750	BY SIMILARITY.
FT	DISULFID	752	761	BY SIMILARITY.
FT	DISULFID	768	779	BY SIMILARITY.
FT	DISULFID	773	788	BY SIMILARITY.
FT	DISULFID	790	799	BY SIMILARITY.
FT	DISULFID	807	818	BY SIMILARITY.

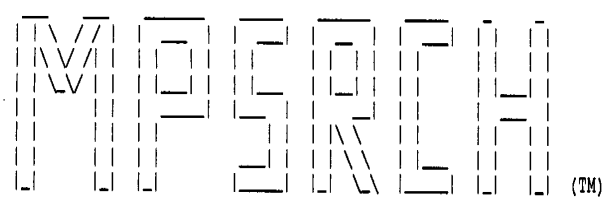
```
FT   DISULFID    812    827       BY SIMILARITY.
FT   DISULFID    829    838       BY SIMILARITY.
FT   DISULFID    845    856       BY SIMILARITY.
FT   DISULFID    850    865       BY SIMILARITY.
FT   DISULFID    867    876       BY SIMILARITY.
FT   DISULFID    882    903       BY SIMILARITY.
FT   DISULFID    897    912       BY SIMILARITY.
FT   DISULFID    914    923       BY SIMILARITY.
FT   DISULFID    930    941       BY SIMILARITY.
FT   DISULFID    935    950       BY SIMILARITY.
FT   DISULFID    952    961       BY SIMILARITY.
FT   DISULFID    968    979       BY SIMILARITY.
FT   DISULFID    973    988       BY SIMILARITY.
FT   DISULFID    990    999       BY SIMILARITY.
FT   DISULFID   1006   1019       BY SIMILARITY.
FT   DISULFID   1011   1028       BY SIMILARITY.
FT   DISULFID   1030   1039       BY SIMILARITY.
FT   DISULFID   1046   1057       BY SIMILARITY.
FT   DISULFID   1051   1069       BY SIMILARITY.
FT   DISULFID   1071   1080       BY SIMILARITY.
FT   DISULFID   1087   1098       BY SIMILARITY.
FT   DISULFID   1092   1110       BY SIMILARITY.
FT   DISULFID   1112   1121       BY SIMILARITY.
FT   DISULFID   1130   1142       BY SIMILARITY.
FT   DISULFID   1136   1155       BY SIMILARITY.
FT   DISULFID   1157   1166       BY SIMILARITY.
FT   CARBOHYD    711    711       POTENTIAL.
FT   CARBOHYD    960    960       POTENTIAL.
FT   CARBOHYD   1139   1139       POTENTIAL.
FT   CONFLICT     43     43       Q -> R (IN REF. 3).
FT   CONFLICT    298    298       L -> P (IN REF. 3).
FT   CONFLICT    884    884       M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match           23.4%;  Score 272;  DB 1;  Length 1964;
  Best Local Similarity  38.7%;  Pred. No. 3.45e-38;
  Matches    36;  Conservative   18;  Mismatches  36;  Indels   3;  Gaps   2;

Db    458 CLCLPGYTGSRCEADHNECLSQPCHPGSTCLDLLATFHCLCPPGLEGRLCEVE-V--NEC 514
          | |: ||||  |  ::::|  : |: |: |:| : :: |||  |  |:|||:     : |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    515 TSNPCLNQAACHDLLNGFQCLCLPGFTGARCEK 547
            :  | | | | |  :   | |||| |: |||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEK 101


Search completed: Fri May 28 09:20:39 1999
Job time : 22 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:20:58 1999; MasPar time 12.73 Seconds
660.074 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-10
Description: (1-154) from US09191647.pep
Perfect Score: 1160
Sequence: 1 DPLPVHHRCECMLGYTGDNC.....EDNGILLYNGDNDHIAVELY 154

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 39.710; Variance 79.104; scale 0.502

[redacted] Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1058	91.2	1531	11	O88279	MEGF4.	1.61e-191
2	625	53.9	1523	11	O88280	MEGF5.	1.96e-101
3	611	52.7	739	4	O75094	MEGF5 (FRAGMENT).	1.44e-98
4	448	38.6	530	5	Q24526	SLIT LOCUS ENCODING A	1.42e-65
5	443	38.2	601	5	Q20204	F40E10.4 PROTEIN (FRAG	1.42e-64
6	333	28.7	721	13	Q91902	X-DELTA-1.	6.30e-43
7	326	28.1	529	5	Q25058	FIBROPELLIN IA (FRAGME	1.42e-41
8	325	28.0	728	13	Q90656	TRANSMEMBRANE PROTEIN	2.21e-41
9	320	27.6	406	5	Q25059	FIBROPELLIN III (FRAGM	2.03e-40
10	313	27.0	717	13	P87357	DELTAD TRANSMEMBRANE P	4.51e-39
11	313	27.0	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	4.51e-39
12	308	26.6	802	13	O57462	DELTAA.	4.10e-38
13	300	25.9	2408	4	Q92566	MYELOBLAST KIAA0279 (F	1.39e-36
14	299	25.8	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	2.15e-36
15	299	25.8	2470	11	O35516	CELL SURFACE PROTEIN.	2.15e-36
16	299	25.8	2531	5	O16004	NOTCH HOMOLOG.	2.15e-36
17	296	25.5	752	13	O42374	NOTCH RECEPTOR PROTEIN	8.02e-36
18	295	25.4	2352	5	O61240	HRNOTCH PROTEIN.	1.24e-35
19	294	25.3	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	1.93e-35
20	292	25.2	1218	4	O14902	TRANSMEMBRANE PROTEIN	4.63e-35
21	292	25.2	1218	4	Q15816	TRANSMEMBRANE PROTEIN	4.63e-35
22	292	25.2	1219	11	Q63722	JAGGED PROTEIN.	4.63e-35
23	292	25.2	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	4.63e-35
24	291	25.1	2447	13	O13149	NOTCH 2 (FRAGMENT).	7.17e-35
25	290	25.0	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	1.11e-34
26	290	25.0	1218	4	O15122	JAGGED1.	1.11e-34
27	285	24.6	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	9.85e-34
28	284	24.5	3313	11	O88278	MEGF2.	1.52e-33
29	283	24.4	1372	5	P91526	SIMILARITY TO MULTIPLE	2.35e-33
30	277	23.9	615	13	O57409	DELTAB.	3.20e-32
31	273	23.5	642	13	P79941	NOTCH LIGAND X-DELTA-2	1.81e-31
32	272	23.4	1964	11	O35442	NOTCH4.	2.80e-31
33	270	23.3	156	5	Q26661	EPIDERMAL GROWTH FACTO	6.65e-31
34	270	23.3	955	4	Q99466	NOTCH4 (FRAGMENT).	6.65e-31
35	270	23.3	1999	4	Q99940	NOTCH4.	6.65e-31
36	270	23.3	2003	4	O00306	NOTCH4.	6.65e-31
37	265	22.8	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	5.76e-30
38	262	22.6	1476	13	Q90285	PUTATIVE EXTRACELLULAR	2.10e-29
39	261	22.5	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	3.22e-29
40	248	21.4	434	11	O55139	JAGGED2 PROTEIN (FRAGM	8.44e-27
41	248	21.4	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	8.44e-27
42	246	21.2	1202	11	P97607	JAGGED2 (FRAGMENT).	1.98e-26
43	245	21.1	585	11	O35675	M-DELTA-LIKE 3 GENE PR	3.03e-26
44	245	21.1	592	11	O88516	DELTA-LIKE 3 ALTERNATE	3.03e-26
45	244	21.0	1827	5	Q20535	SIMILARITY TO EGF-TYPE	4.63e-26

ALIGNMENTS

RESULT 1
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 91.2%; Score 1058; DB 11; Length 1531;
Best Local Similarity 95.2%; Pred. No. 1.61e-191;
Matches 140; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Db  1068 RCECVPGYTGDNCSKNQDDCKDHQCQNGAQCVDEINSYACLCAEGYSGQLCEIPPAPRNS 1127
             ||||: |||||||| |||||||| ||||||||||:||||||| |||||||||||||:|
Qy       8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db  1128 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANI 1187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANI 127

Db  1188 TLQVSTAEDNGILLYNGDNDHIAVELY 1214
             |||||||||||||||||||||||||
Qy     128 TLQVSTAEDNGILLYNGDNDHIAVELY 154
```

RESULT 2

```
ID   O88280      PRELIMINARY;      PRT;  1523 AA.
AC   O88280;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5.
GN   MEGF5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011531; D1033424; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1523 AA;  167767 MW;  2BD845D0 CRC32;

  Query Match           53.9%;  Score 625;  DB 11;  Length 1523;
  Best Local Similarity  54.3%;  Pred. No. 1.96e-101;
  Matches   82;  Conservative   33;  Mismatches  32;  Indels   4;  Gaps   3;

Db   1059 RCECVPGYSGKLCETDNDDCVAHKCRHGAQCVDAVNGYTCICPQGFSGLFCEHPPPMVLL 1118
            ||||: ||:|  |  : |||  |||::|||||| ||:|:|:| :|:|| :|| ||:
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPA-P-- 64

Db   1119 QTSPCDQYECQNGAQCIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRP 1178
            ::| |:  |||||| |:     | |:| |||:|| ||||::|||| :|:|:::    :|
Qy     65 RSS-CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWP 123

Db   1179 QANISLQVATDKDNGILLYKGDNDPLALELY 1209
            :|||:|||:|  |||||||:|||| :|:|||
Qy    124 RANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT   3
ID   O75094      PRELIMINARY;      PRT;  739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;

  Query Match           52.7%;  Score 611;  DB 4;  Length 739;
  Best Local Similarity  52.2%;  Pred. No. 1.44e-98;
  Matches   82;  Conservative   33;  Mismatches  38;  Indels   4;  Gaps   3;

Db   269 PLDKGFSCECVPGYSGKLCETDNDDCVAHKCRHGAQCVDTINGYTCTCPQGFSGPFCEHP 328
            ||     |||: ||:|  |  : |||  |||::|||||| :|:|:| | :|:|| :|| |
Qy     2 PLPVHHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIP 61

Db   329 PPMVLLQTSPCDQYECQNGAQCIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELA 388
            |:    ::| |:  |||||| |:     | |:| |||:|| ||||::|||| :|:|::::
Qy    62 PA-P--RSS-CEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db   389 SAKVRPQANISLQVATDKDNGILLYKGDNDPLALELY 425
               :|:|||:|||:|  |||||||:|||| :|:|||
Qy   118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT   4
ID   Q24526      PRELIMINARY;      PRT;  530 AA.
AC   Q24526;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SLIT LOCUS ENCODING A PROTEIN ASSOCIATED WITH NEURAL DEVELOPMENT WITH
DE   52D EGF HOMOLOGOUS DOMAINS (FRAGMENT).
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 89077533.
RA   ROTHBERG J.M., HARTLEY D.A., WALTHER Z., ARTAVANIS-TSAKONAS S.;
RT   "slit: an EGF-homologous locus of D. melanogaster involved in the
RT   development of the embryonic central nervous system.";
RL   CELL 55:1047-1059(1988).
DR   EMBL; M23543; G514357; -.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
KW   NEUROGENESIS; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     530    530
SQ   SEQUENCE   530 AA;  59457 MW;  10E5764D CRC32;

  Query Match           38.6%;  Score 448;  DB 5;  Length 530;
  Best Local Similarity  41.4%;  Pred. No. 1.42e-65;
  Matches   65;  Conservative   39;  Mismatches  44;  Indels   9;  Gaps   5;

Db   167 HYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMIS 226
            |  |:|  |: | ||::| ||| :|:|||||: ||| :| | | | : |:|  ||    |:
Qy     6 HHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCE----IP 61

Db   227 MMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELE 286
               |::| |:  || :|       |||  :|:| ||: |  || | |::||  ::::::
Qy    62 PA-PRSS-CEGTECQNGA--NCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 117

Db   287 PLRTRPEANVTI-VFSSGQNGILMYDGQDAHLAVELF 322
            |: :| ||:|: | :: :||||:|:|:: |:||||:
Qy   118 DLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT   5
ID   Q20204      PRELIMINARY;      PRT;  601 AA.
AC   Q20204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F40E10.4 PROTEIN (FRAGMENT).
GN   F40E10.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   SMYE R.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RL   EMBL; Z69792; E1346469; -.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match           38.2%;  Score 443;  DB 5;  Length 601;
  Best Local Similarity  42.5%;  Pred. No. 1.42e-64;
  Matches    68;  Conservative   27;  Mismatches  56;  Indels   9;  Gaps   8;

Db    195 PINGSYSCMCSPGFTGNNCETNIDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIP 254
          |:    | |  |:||:||  | ||||:  ||||: ||| : || |||  ||:|| ||||
Qy      2 PLPVHHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIP 61

Db    255 PMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYL 314
          |     : |: || :  | ::||  | |    | |  ||:|| |:: :|| | :  :||
Qy     62 PAPRSSCEGTE-CQ-N--G-ANCVD-QGSRP-VCQCLPGFGGPECEKLLSVNFVDRDTYL 114

Db    315 ALDPLAS-D-GTITMTLRTTSKIGILLYYGDDHFVSAELY 352
           : | :   : ||: : |:   ||||| ||:  :: |||
Qy    115 QFTDLQNWPRANITLQVSTAEDNGILLYNGDNDHIAVELY 154


RESULT    6
ID   Q91902      PRELIMINARY;      PRT;   721 AA.
AC   Q91902;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   X-DELTA-1.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95319507.
RA   HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT   "Expression of a Delta homologue in prospective neurons in the
RT   chick.";
RL   NATURE 375:787-790(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95319503.
RA   CHITNIS A.B., HENRIQUE D., LEWIS J., ISH-HOROWICZ D., KINTNER C.R.;
RT   "Primary neurogenesis in Xenopus embryos regulated by a homologue of
RT   the Drosophila neurogenic gene Delta.";
RL   NATURE 375:761-766(1995).
DR   EMBL; L42229; G807696; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   721 AA;  79922 MW;  028040EF CRC32;

  Query Match           28.7%;  Score 333;  DB 13;  Length 721;
  Best Local Similarity  42.3%;  Pred. No. 6.30e-43;
  Matches    44;  Conservative   18;  Mismatches  39;  Indels   3;  Gaps   2;

Db    423 EDLGNSYICQCCQEGFSGRNCDDNLDDCTSFPCQNGGTCQDGINDYSCTCPPGYIGKNCSM 482
          : |     |:|  |::| || :| |||    ||||: | | :| |:| |  || |  | :
Qy      1 DPLPVHHRCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEI 60

Db    483 P-ITK--CEHNPCHNGATCHERNNRYVCQCARGYGGNNCQFLLP 523
          |   :  ||   |:||| | :: :| ||||  |:|| :|: ||:
Qy     61 PPAPRSSCEGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT    7
ID   Q25058      PRELIMINARY;      PRT;   529 AA.
AC   Q25058;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN IA (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33861; G499686; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 10.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   529 AA;  55543 MW;  6385F322 CRC32;

  Query Match           28.1%;  Score 326;  DB 5;  Length 529;
  Best Local Similarity  44.6%;  Pred. No. 1.42e-41;
  Matches    41;  Conservative   18;  Mismatches  30;  Indels   3;  Gaps   2;

Db    273 CSCVQGFTGSDCETNINECASGPCQNGGTCVDGVNGFVCQCPPNYTGTYCEIS-L--DAC 329
          | |: |:|| :|  | ::|    ||||: ||| ||:: | |   |:|  |||:     :|
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    330 SSMPCQNGATCVNVGANYICECPPGFAGQNCE 361
          :  ||||| ||: |:  :|:| |||:| :||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT    8
ID   Q90656      PRELIMINARY;      PRT;   728 AA.
AC   Q90656;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE PROTEIN C-DELTA-1.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 95319507.
RA   HENRIQUE D., ADAM J., MYAT A., CHITNIS A., LEWIS J., ISH-HOROWICZ D.;
RT   "Expression of a Delta homologue in prospective neurons in the
RT   chick.";
RL   NATURE 375:787-790(1995).
DR   EMBL; U26590; G882412; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   728 AA;  79861 MW;  7439F575 CRC32;
```

```
  Query Match          28.0%;  Score 325;  DB 13;  Length 728;
  Best Local Similarity  44.8%;  Pred. No. 2.21e-41;
  Matches   43;  Conservative   18;  Mismatches  32;  Indels   3;  Gaps   2;

Db    436 CQCQAGFTGRHCDDNVDDCASFPCVNGGTCQDGVNDYSCTCPPGYNGKNCSTP-VSR--C 492
          |:|  |:|| :| :| |||    | ||: | | || |:| |  ||:|  |  |  :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    493 EHNPCHNGATCHERSNRYVCECARGYGGLNCQFLLP 528
          |   |:||| | ::::| ||:|  |:|| :|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT   9
ID   Q25059      PRELIMINARY;      PRT;   406 AA.
AC   Q25059;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN III (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33862; G499688; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 7.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   406 AA;  43475 MW;  45E6EE2C CRC32;

  Query Match          27.6%;  Score 320;  DB 5;  Length 406;
  Best Local Similarity  43.5%;  Pred. No. 2.03e-40;
  Matches   40;  Conservative   19;  Mismatches  30;  Indels   3;  Gaps   2;

Db     74 CNCIPGFDGDNCENNINECASNPCQNGGVCIDGVNGFVCTCQPGYTGTLCETD-I--DEC 130
          |:|: |: |||| :| ::|  : ||||: |:| ||:: | |  ||:| |||       |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    131 ASNPCQNGGVCTDLVNMYTCDCLAGFTGSNCE 162
          :  ||||: | |  :   |:||:|| |::||
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


SULT   10
     P87357      PRELIMINARY;      PRT;   717 AA.
     P87357;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAD TRANSMEMBRANE PROTEIN PRECURSOR.
GN   DELTAD.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97346722.
RA   DORNSEIFER P., TAKKE C., CAMPOS-ORTEGA J.A.;
RT   "Overexpression of a zebrafish homologue of the Drosophila neurogenic
RT   gene Delta perturbs differentiation of primary neurons and somite
RT   development.";
RL   MECH. DEV. 63:159-171(1997).
DR   EMBL; Y11760; E307461; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   SIGNAL; TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    717       DELTAD TRANSMEMBRANE PROTEIN.
SQ   SEQUENCE   717 AA;  79061 MW;  5CC32ECA CRC32;

  Query Match          27.0%;  Score 313;  DB 13;  Length 717;
  Best Local Similarity  40.8%;  Pred. No. 4.51e-39;
  Matches   40;  Conservative   21;  Mismatches  34;  Indels   3;  Gaps   2;

Db    427 CQCPEGFTGTHCEDNIDECATYPCQNGGTCQDGLSDYTCTCPPGYTGKNC-TSAVNK--C 483
          |:|  |:|| :| :| |:|    ||||: | | :: |:| |  ||:|  |  ::  :  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    484 LHNPCHNGATCHEMDNRYVCACIPGYGGRNCQFLLPEN 521
            |:||| | :  :| || |:||:|| :|: ||: |
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVN 106


RESULT   11
ID   Q19350      PRELIMINARY;      PRT;  1722 AA.
AC   Q19350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO EGF-LIKE REPEATS. NCBI GI: 1125776.
GN   F11C7.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A., VETTER J.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U42839; G1125776; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PFAM; PF00008; EGF; 24.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1722 AA;  188383 MW;  CCFB86B8 CRC32;

  Query Match          27.0%;  Score 313;  DB 5;  Length 1722;
  Best Local Similarity  41.9%;  Pred. No. 4.51e-39;
  Matches   39;  Conservative   18;  Mismatches  34;  Indels   2;  Gaps   2;

Db    148 KCACPPGFVGDHCETDEDECKENFCQNGADCENLKGSYECKCLKGFSGKYCEIQ-D-KKQ 205
          :| |  |: ||:|  ::|:||:: |||||:| :   || | |: |:||  |||    :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    206 CTSDYCHNNGQCISTGSDLSCKCSPGFDGAFCE 238
          | :  |:| : |:  ||   | | ||| |: ||
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   12
```

```
ID   O57462      PRELIMINARY;      PRT;   802 AA.
AC   O57462;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAA.
GN   DELTAA.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98165392.
RA   APPEL B., EISEN J.S.;
RT   "Regulation of neuronal specification in the zebrafish spinal cord by
RT   Delta function.";
RL   DEVELOPMENT 125:371-380(1998).
RA   EMBL; AF030031; G2809389; -.
DR   PROSITE; PS01186; EGF_2; 8.
KW   GLYCOPROTEIN.
SQ   SEQUENCE   802 AA;  88941 MW;  42F041BD CRC32;

  Query Match          26.6%;  Score 308;  DB 13;  Length 802;
  Best Local Similarity  40.6%;  Pred. No. 4.10e-38;
  Matches   39;  Conservative   22;  Mismatches  32;  Indels   3;  Gaps   2;

Db    469 CQCPDGFTGMNCDRAGDECSMYPCQNGGTCQEGASGYMCTCPPGYTGRNCS-SPVSR--C 525
          |:|  |:|| ||    |:|    ||||: | :  ::| | |  ||:|: |  :| :|  |
Qy      9 CECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSSC 68

Db    526 QHNPCHNGATCHERNNRYVCACVSGYGGRNCQFLLP 561
          :   |:||| | :: :| || |::|:|| :|: ||:
Qy     69 EGTECQNGANCVDQGSRPVCQCLPGFGGPECEKLLS 104


RESULT   13
ID   Q92566      PRELIMINARY;      PRT;  2408 AA.
AC   Q92566;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELOBLAST KIAA0279 (FRAGMENT).
GN   KIAA0279.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA RES. 3:321-329(1996).
DR   EMBL; D87469; D1014097; -.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PFAM; PF00008; EGF; 6.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF00054; laminin_G; 1.
KW   GLYCOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   2408 AA;  261740 MW;  CDBA2001 CRC32;

  Query Match          25.9%;  Score 300;  DB 4;  Length 2408;
  Best Local Similarity  33.1%;  Pred. No. 1.39e-36;
  Matches   46;  Conservative   33;  Mismatches  55;  Indels   5;  Gaps   5;

Db    758 RCRCPPGFTGDYCETEVDLCYSRPCGPHGRCRSREGGYTCLCRDGYTGEHCEVS-ARSGR 816
          || |  |:||| |  : | |  : |   ::|     :|:||| :||:|: ||:: |  :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    817 CTPGVCKNGGTCVNLLVGGFKCDCPSGDFEKPYCQ-VTTRSFPAH-SFITFRGLRQRFHF 874
          |    | ||: ||:       |:| :| |  | |: : : :|  : ::: |  |: : :
Qy     68 CEGTECQNGANCVDQGSRPV-CQCLPG-FGGPECEKLLSVNFVDRDTYLQFTDLQNWPRA 125

Db    875 TLALSFATKERDGLLLYNG 893
          ::|  :| | :|:|||||
Qy    126 NITLQVSTAEDNGILLYNG 144


RESULT   14
ID   Q06008      PRELIMINARY;      PRT;  1203 AA.
AC   Q06008;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH PROTEIN HOMOLOG 2 (MOTCH B PROTEIN) (FRAGMENT).
GN   NOTCH2 OR MOTCH B.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
DR   EMBL; X68279; G287990; -.
DR   MGD; MGI:97364; NOTCH2.
DR   PFAM; PF00008; EGF; 27.
DR   PFAM; PF00066; notch; 1.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT.
FT   NON_TER       1      1
FT   NON_TER    1203   1203
SQ   SEQUENCE   1203 AA;  128982 MW;  A5A95551 CRC32;

  Query Match          25.8%;  Score 299;  DB 11;  Length 1203;
  Best Local Similarity  38.7%;  Pred. No. 2.15e-36;
  Matches   36;  Conservative   20;  Mismatches  34;  Indels   3;  Gaps   2;

Db    124 HCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFKGVHCELE-V--NE 180
          :|||: ||:|  |  : ::|    ||| | |:| : :::|||: |: |  ||:     :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    181 CQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQ 213
          |::  | | : |||  :|  | | ||| || |:
Qy     68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


RESULT   15
ID   O35516      PRELIMINARY;      PRT;  2470 AA.
AC   O35516;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CELL SURFACE PROTEIN.
GN   NOTCH2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
```

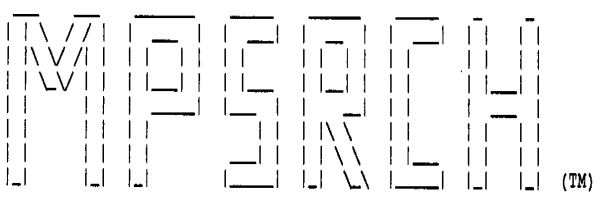
```
RA   HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL   SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D32210; D1022953; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match           25.8%;  Score 299;  DB 11;  Length 2470;
  Best Local Similarity  38.7%;  Pred. No. 2.15e-36;
  Matches    36;  Conservative   20;  Mismatches  34;  Indels   3;  Gaps   2;

Db    439 HCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFKGVHCELE-V--NE 495
          :|||: ||:|  |  : ::|    ||| | |:| : :::|||: |: |  ||:     :
Qy      8 RCECMLGYTGDNCSENQDDCKDHKCQNGAQCVDEVNSYACLCVEGYSGQLCEIPPAPRSS 67

Db    496 CQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQ 528
          |::  | | : |||  :|  | | ||| || |:
       68 CEGTECQNGANCVDQGSRPVCQCLPGFGGPECE 100


Search completed: Fri May 28 09:21:48 1999
Job time : 50 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:23:41 1999; MasPar time 6.79 Seconds
344.712 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-11
Description: (1-110) from US09191647.pep
Perfect Score: 873
Sequence: 1 AFKCHHGQCHISDRGEPYCL.....GSSFVEEVERHLECGCRACS 110

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 28.127; Variance 120.970; scale 0.233

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	442	50.6	1534	30	W46966	Amino acid sequence o	1.90e-31
2	167	19.1	1480	5	R25079	Drosophila SLIT prote	5.59e-06
3	131	15.0	1872	36	W68510	Partial human Notch-3	6.56e-03
4	131	15.0	2321	36	W49698	Human Notch3 protein.	6.56e-03
5	129	14.8	374	22	W07663	Human transforming gr	9.62e-03
6	129	14.8	374	27	W37497	Human TMP-2.	9.62e-03
7	126	14.4	379	5	R25565	Beta-IG-M1.	1.71e-02
8	121	13.9	1055	29	W44298	Human serrate 2 prote	4.40e-02
9	121	13.9	1212	29	W44299	Human serrate 2.	4.40e-02
10	121	13.9	1257	19	W05834	Human Serrate-2 (HJ2)	4.40e-02
11	120	13.7	612	28	W39256	Human partial mature	5.31e-02
12	120	13.7	737	28	W39257	Human membrane protei	5.31e-02
13	118	13.5	121	12	R64230	HCT-15 contg. hTG-700	7.73e-02
14	114	13.1	381	26	W35957	Human monocyte mature	1.63e-01
15	114	13.1	381	26	W35730	Human cysteine rich p	1.63e-01
16	114	13.1	727	21	W11719	C-Delta-1 polypeptide	1.63e-01
17	114	13.1	740	21	W00876	C-Delta-1 polypeptide	1.63e-01
18	113	12.9	177	8	R40167	Recombinant growth fa	1.96e-01
19	112	12.8	44	3	R15351	Tumour cell growth in	2.36e-01
20	112	12.8	46	3	R15350	Tumour cell growth in	2.36e-01
21	112	12.8	157	21	W11730	H-Delta-1 polypeptide	2.36e-01
22	112	12.8	520	25	W18348	Proliferation and dif	2.36e-01
23	112	12.8	660	21	W11725	H-Delta-1 polypeptide	2.36e-01
24	112	12.8	702	25	W18349	Proliferation and dif	2.36e-01
25	112	12.8	723	25	W18353	Proliferation and dif	2.36e-01
26	111	12.7	183	30	W46968	Amino acid sequence o	2.85e-01
27	111	12.7	228	30	W46967	Amino acid sequence o	2.85e-01
28	109	12.5	44	7	R38212	Tumour cell prolifera	4.11e-01
29	109	12.5	46	7	R41537	Tumour cell prolifera	4.11e-01
30	109	12.5	46	9	R45226	P-1 cancer cell growt	4.11e-01
31	109	12.5	51	3	P50547	Protein analogue enco	4.11e-01
32	109	12.5	179	37	W75100	Human secreted protei	4.11e-01
33	108	12.4	383	10	R56166	Neuroendocrine tumor	4.94e-01
34	107	12.3	385	10	R56167	Neuroendocrine tumor	5.94e-01
35	106	12.1	50	2	P98501	Analouge of native hu	7.13e-01
36	106	12.1	77	16	R80786	Human heparin binding	7.13e-01
37	106	12.1	82	16	R92917	Met-Cys-Ala-Met-Ala-H	7.13e-01
38	106	12.1	160	1	P91000	Transforming growth f	7.13e-01
39	106	12.1	208	16	R80785	Human precursor hepar	7.13e-01
40	106	12.1	208	16	R80787	Monkey precursor hepa	7.13e-01
41	106	12.1	208	16	R92897	Human HBEGF precursor	7.13e-01
42	106	12.1	208	4	R23998	EGF/HB-EHM.	7.13e-01
43	106	12.1	208	13	R66190	Diphtheria toxin (DT)	7.13e-01
44	106	12.1	333	16	R92915	HBEGF-linker-SAP.	7.13e-01
45	106	12.1	333	16	R92900	HBEGF Val Met saporin	7.13e-01

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 50.6%; Score 442; DB 30; Length 1534;
Best Local Similarity 47.7%; Pred. No. 1.90e-31;
Matches 53; Conservative 24; Mismatches 33; Indels 1; Gaps 1;

Db 1424 glqclhghcqasgtkgahcvcdpgfsgelceqesecrgdpvrdfhqvqrgyaicqttrpl 1483
:: | ||:|: | : |:|:||||| ||||: | |: ||: : |: || | |: :
Qy 1 AFKCHHGQCHISDRGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASKV 60

Db 1484 swvecrgscpgqgccqglrlkrrkftfecsdqtsfaeevekptkcgcalca 1534
: :||||:| | ||| :| ||||: |:|:||:|| ||||: ||| |:

```
Qy     61 PIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRACS 110


RESULT   2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;
```

```
  Query Match          19.1%;  Score 167;  DB 5;  Length 1480;
  Best Local Similarity  29.4%;  Pred. No. 5.59e-06;
  Matches   35;  Conservative   25;  Mismatches  47;  Indels  12;  Gaps  12;

Db   1361 kcrrgsrcvpnsnardgyqckckhgqrgrycdqgegstepptvtaastcrkeqvreyyte 1420
          ||::| :|   |:  : | | |  |  |: |:| | :     | |   |: | ::| :
Qy      3 KCHHG-QC-HISDRGEPY-CLCQPGFSGHHCEQ-ENPCMGEIVREA-I-RR-Q-KDY-AS 53

Db   1421 ndcrsrqplkyakcvggcgnqccaakivrrrkvrmvcsnnrkyiknldivrkcgctkkc 1479
               |: |::   | |||| || :   :|||    |::   :: :::    ||| : |
Qy     54 CATASKVPIM-E-CRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGC-RAC 109


RESULT   3
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 328
FT                   /note= "encoded by NAN"
FT   Misc_difference 401
FT                   /note= "encoded by GNN"
FT   Misc_difference 403
FT                   /note= "encoded by GNC"
FT   Misc_difference 406
FT                   /note= "encoded by GNN"
FT   Misc_difference 409
FT                   /note= "encoded by NNT"
FT   Misc_difference 420
FT                   /note= "encoded by GNC"
FT   Misc_difference 706
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
SQ   Sequence   1872 AA;
```

```
  Query Match          15.0%;  Score 131;  DB 36;  Length 1872;
  Best Local Similarity  42.1%;  Pred. No. 6.56e-03;
  Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db    989 cqagggc-vdedsshycvcpegrtgshceqevdpclaq 1025
          |: | || : : :  ||:|  | :| ||||| :||::: 
Qy      4 CHHG-QCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT    4
ID   W49698 standard; Protein; 2321 AA.
AC   W49698;
DT   21-DEC-1998  (first entry)
DE   Human Notch3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy.
OS   Homo sapiens.
PN   FR2751986-A1.
PD   06-FEB-1998.
PF   16-APR-1997; 004680.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133138/13.
DR   N-PSDB; V57001.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1.1-1.8; 45pp; French.
CC   This sequence represents the human Notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of
CC   the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC   and leukoencephalopathy (CADASIL) type.  Blocking expression of a
CC   mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC   gene or protein can be used to treat CADASIL or related disorders.
SQ   Sequence   2321 AA;

  Query Match          15.0%;  Score 131;  DB 36;  Length 2321;
  Best Local Similarity  42.1%;  Pred. No. 6.56e-03;
  Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db   1055 cqagggc-vdedsshycvcpegrtgshceqevdpclaq 1091
          |: | || : : :  ||:|  | :| ||||| :||::: 
Qy      4 CHHG-QCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT    5
ID   W07663 standard; Protein; 374 AA.
AC   W07663;
DT   30-JUL-1997  (first entry)
DE   Human transforming growth factor alpha HII .
KW   TGFalpha-HII; AIDS; dementia; ocular disease; kidney disorder;
KW   liver disorder; hair follicle development; angiogenesis; ulcer;
KW   corneal incision; embryogenesis; gene therapy; neoplasia;
KW   psoriasis.
OS   Homo sapiens.
PN   WO9636709-A1.
PD   21-NOV-1996.
PF   19-MAY-1995; U06386.
PR   19-MAY-1995; WO-U06386.
PR   12-JUN-1995; ZA-004848.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Meissner PS,  Ni J,  Wei Y;
DR   WPI; 97-012084/01.
DR   N-PSDB; T45428.
PT   Nucleic acid encoding human transforming growth factor-alpha HII -
PT   useful for treating, e.g. ocular diseases, kidney and liver
PT   disorders, or to stimulate wound healing etc
PS   Claim 19; Page 48-49; 73pp; English.
CC   The present sequence represents the 374 amino acid sequence for
CC   human transforming growth factor alpha HII (TGFalpha-HII). Human
CC   TGFalpha-HII can be used to stimulate wound healing; to restore normal
CC   neurological functioning after trauma or AIDS, dementia; to treat ocular
CC   diseases, kidney and liver disorders; promote hair follicle development;
CC   to stimulate angiogenesis for treating burns, ulcers, corneal incisions;
CC   and to stimulate embryogenesis. The TGFalpha-HII can be used directly or
CC   is generated in situ, i.e. by gene therapy. Antagonists of TGFalpha-HII
CC   are useful for treating neoplasia and for treating certain skin
CC   disorders, such as psoriasis. Detecting mutations in the polynucleotide
CC   sequence is used for diagnosing diseases (or susceptibility to diseases)
CC   which result from underexpression of TGFalpha-HII.
SQ   Sequence   374 AA;

  Query Match          14.8%;  Score 129;  DB 22;  Length 374;
  Best Local Similarity  31.0%;  Pred. No. 9.62e-03;
  Matches   27;  Conservative   19;  Mismatches   35;  Indels   6;  Gaps   6;

Db    271 gf-cmhgkcehsinmqepscrcdagytgqhcekkdysvlyvvpgpvrfqyvlia-avigt 328
          :| | || | |  :  || | |::|::|:|||  :  :  :|  ::| |    : |  :
Qy      1 AFKCHHGQC-HISDRGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASK 59

Db    329 iqiavicvvvlcitrkcprsnrihrqk 355
          : | : |    | |  | :: |  |:|
Qy     60 VPI-MECRGG-CGT-TCCQPIRSKRRK 83


RESULT    6
ID   W37497 standard; Protein; 374 AA.
AC   W37497;
DT   20-APR-1998  (first entry)
DE   Human TMP-2.
KW   Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW   brain specific nucleosome assembly protein; diagnosis; therapy;
KW   skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
KW   nel-related protein type 1; nel-related type 2; hereditary disease;
KW   cancer.
OS   Homo sapiens.
PN   EP-796913-A2.
PD   24-SEP-1997.
PF   19-MAR-1997; 104842.
PR   05-MAR-1997; JP-069163.
PR   19-MAR-1996; JP-063410.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Fujiwara T, Horie M, Watanabe T;
DR   WPI; 97-459830/43.
DR   N-PSDB; V01874, V01875.
PT   Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT   useful for diagnosis or therapy of hereditary disease and cancer
PS   Claim 10; Page 67-68; 123pp; English.
CC   The present sequence represents a TMP-2 isolated from a human foetal
CC   brain cDNA library. The nucleotide or amino acid sequences are useful
CC   for in-vitro diagnosis of hereditary diseases and cancer and for
CC   preparation of pharmaceuticals.
SQ   Sequence   374 AA;

  Query Match          14.8%;  Score 129;  DB 27;  Length 374;
  Best Local Similarity  31.0%;  Pred. No. 9.62e-03;
  Matches   27;  Conservative   19;  Mismatches   35;  Indels   6;  Gaps   6;

Db    271 gf-cmhgkcehsinmqepscrcdagytgqhcekkdysvlyvvpgpvrfqyvlia-avigt 328
          :| | || | |  :  || | |::|::|:|||  :  :  :|  ::| |    : |  :
Qy      1 AFKCHHGQC-HISDRGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASK 59

Db    329 iqiavicvvvlcitrkcprsnrihrqk 355
          : | : |    | |  | :: |  |:|
Qy     60 VPI-MECRGG-CGT-TCCQPIRSKRRK 83


RESULT    7
ID   R25565 standard; Protein; 379 AA.
AC   R25565;
DT   18-JAN-1993  (first entry)
```

DE Beta-IG-M1.
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW embryo; fibroblasts; TGF-beta.
OS Mus musculus.
PN EP-495674-A.
PD 22-JUL-1992.
PF 17-JAN-1992; 300429.
PR 18-JAN-1991; US-642991.
PR 10-JAN-1992; US-816270.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR WPI; 92-243508/30.
DR N-PSDB; Q26421.
PT TGF-beta induced gene family - encodes proteins involved in
PT growth and differentiation effects of TGF-beta-1
PS Claim 2; Fig 1; 35pp; English.
CC The protein sequence was deduced from the DNA sequence obtd. by
CC screening a cDNA library made from AKR-2B mouse cells induced with
CC TGF-beta1 and cyclohexamide with two probes from untreated AKR-2B
CC mRNA and AKR-2B mRNA from cells treated with cyclohexamide and TGF-
CC beta1. The proteins encoded by hybridising colonies (beta-IG-M1 and
CC beta-IG-M2) contain 38 Cys residues and are induced by TGF-beta1.
CC Beta-IG-M1 displays 80 percent homology to the CEF-10 protein
CC induced by v-src in chicken embryo fibroblasts and is identical
CC to the protein encoded by cyr61, an immediate early response gene
CC induced in quiescent BALB 3T3 cells by serum treatment. Residues
CC 49-56 of beta-IG-M1 conform to the GCGCCXXC motif reported in the
CC amino half of insulin-like growth factor (IGF) binding proteins.
CC The C-terminal Cys rich region of beta-IG-M1, -M2 and CEF-10 contain
CC an amino acid sequence with strong homology to a motif found near the
CC C-terminal of the malarial circumsporozoite (CS) protein, which is
CC highly conserved among all species of malarial parasites sequenced
CC to date (designated region II). This motif is also found in
CC other proteins which have cell adhesive properties that mediate
CC cell-cell and cell-extracellular matrix interactions, such as
CC properdin, thrombospondin, and TRAP. The proteins encoded by
CC TGF-beta induced genes are likely to be involved in mediation of
CC the biological effects of TGF-beta relating to cell growth and
CC differentiation. See also R25566.
SQ Sequence 379 AA;

Query Match 14.4%; Score 126; DB 5; Length 379;
Best Local Similarity 32.3%; Pred. No. 1.71e-02;
Matches 20; Conservative 12; Mismatches 27; Indels 3; Gaps 3;

Db 298 yagcssvkkyrpkyc-gscvdgrcctplqtrtvkmrfrcedgemfsknvmmiqsckcnyn 356
Qy 51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

357 cp 358
109 CS 110

RESULT 8
ID W44298 standard; Protein; 1055 AA.
AC W44298;
DT 19-JUN-1998 (first entry)
DE Human serrate 2 protein fragment.
KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.
PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15181.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Claim 2; Page 57-62; 103pp; Japanese.
CC The present sequence represents a human serrate 2 protein fragment. The
CC present invention also describes a method for the preparation of the
CC polypeptides, and antibodies binding to the polypeptide and its
CC fragments. The polypeptide and its fragments expressed by the serrate-2
CC gene can be used to inhibit stem (especially blood stem) cell
CC differentiation and to inhibit endothelial cell growth. They may be
CC incorporated in a cell culture media for culturing undifferentiated
CC stem cells. They can also be used for treatment of neoplasms such as
CC leukaemia. The antibodies can be used for the diagnosis of malignant
CC tumours.
SQ Sequence 1055 AA;

Query Match 13.9%; Score 121; DB 29; Length 1055;
Best Local Similarity 34.1%; Pred. No. 4.40e-02;
Matches 15; Conservative 15; Mismatches 10; Indels 4; Gaps 4;

Db 580 cgphgrc-vsqpggnfscicdsgftgtycheniddclgqpcrng 622
Qy 4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44

RESULT 9
ID W44299 standard; Protein; 1212 AA.
AC W44299;
DT 19-JUN-1998 (first entry)
DE Human serrate 2.
KW Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW leukaemia; endothelial cell; tumour.
OS Homo sapiens.
PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15181.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Claim 3; Page 62-68; 103pp; Japanese.
CC The present sequence represents human serrate 2. The present invention
CC also describes a method for the preparation of the polypeptides, and
CC antibodies binding to the polypeptide and its fragments. The polypeptide
CC and its fragments expressed by the serrate-2 gene can be used to inhibit
CC stem (especially blood stem) cell differentiation and to inhibit
CC endothelial cell growth. They may be incorporated in a cell culture
CC media for culturing undifferentiated stem cells. They can also be used
CC for treatment of neoplasms such as leukaemia. The antibodies can be used
CC for the diagnosis of malignant tumours.
SQ Sequence 1212 AA;

Query Match 13.9%; Score 121; DB 29; Length 1212;
Best Local Similarity 34.1%; Pred. No. 4.40e-02;
Matches 15; Conservative 15; Mismatches 10; Indels 4; Gaps 4;

Db 580 cgphgrc-vsqpggnfscicdsgftgtycheniddclgqpcrng 622
Qy 4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44

RESULT 10
ID W05834 standard; Protein; 1257 AA.
AC W05834;
DT 28-JAN-1997 (first entry)
DE Human Serrate-2 (HJ2).
KW Serrate-2; human jagged-2; HJ2; Notch; cell differentiation;
KW cell fate; central nervous system; cancer; tissue repair; therapy;
KW diagnosis; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..912

```
FT                /label= Extracellular_domain
FT                /note= "a deletion in the encoding cDNA clone
FT                results in loss of part of the Serrate-2
FT                signal peptide and beginning of the DSL
FT                domain
FT  domain        26..70
FT                /label= DSL
FT                /note= "region of homology with Drosophila Delta
FT                and Serrate, predicted to mediate binding
FT                with Notch"
FT  domain        75..735
FT                /label= ELR
FT                /note= "epidermal growth factor-like repeat domain"
FT  region        75..105
FT                /label= ELR1
FT  region        106..140
FT                /label= ELR2
FT  region        141..180
FT                /label= ELR3
FT  region        181..218
FT                /label= ELR4
FT  region        219..256
FT                /label= ELR5
FT  region        257..294
FT                /label= ELR6
FT  region        295..331
FT                /label= ELR7
FT  region        332..369
FT                /label= ELR8
FT  region        370..407
FT                /label= ELR9
FT  region        408..435
FT                /label= Partial_ELR
FT  region        436..469
FT                /label= Partial_ELR
FT  region        470..507
FT                /label= ELR10
FT  region        508..545
FT                /label= ELR11
FT  region        546..584
FT                /label= ELR12
FT  region        585..622
FT                /label= ELR13
FT  region        623..660
FT                /label= ELR14
FT  region        664..701
FT                /label= ELR15
FT  region        702..718
FT                /label= Partial_ELR
FT  region        719..735
FT                /label= Partial_ELR
FT  domain        913..933
FT                /label= Transmembrane_domain
FT  domain        934..1257
FT                /label= Intracellular_domain
PN  WO9627610-A1.
PD  12-SEP-1996.
PF  07-MAR-1996; U03172.
PR  07-MAR-1995; US-400159.
PA  (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA  (UYYA ) UNIV YALE.
PI  Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI  Lewis JH, Mann RS, Myat AM;
DR  WPI; 96-425379/42.
DR  N-PSDB; W05834.
PT  Vertebrate Serrate protein and related DNA - used to treat or
PT  prevent malignancies characterised by increased Notch activity.
PS  Claim 5; Page 104-107; 161pp; English.
CC  Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC  for the zygotic neurogenic locus Notch, and are believed to play a
CC  major role in determining cell fates (differentiation) in the
CC  central nervous system. Their amino acid sequences were deduced
CC  from cDNA clones (see also T40090-91) isolated from human foetal
CC  brain cDNA libraries. The proteins, antibodies raised to them,
CC  and encoding nucleic acids can be used in the detection of
CC  Serrate sequences and in the treatment of disorders of cell fate
CC  or differentiation, partic. cancer, nervous system disorders
CC  and in tissue repair or regeneration.
SQ  Sequence  1257 AA;

  Query Match          13.9%;  Score 121;  DB 19;  Length 1257;
  Best Local Similarity  34.1%;  Pred. No. 4.40e-02;
  Matches   15;  Conservative   15;  Mismatches  10;  Indels   4;  Gaps   4;

Db    436 cgphgrc-vsqpggnfscicdsgftgtycheniddclgqpcrng 478
          |  ||:| :|: |  : |:|::||:|  | :: : |:|:  |::
Qy      4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44


RESULT  11
ID  W39256 standard; protein; 612 AA.
AC  W39256;
DT  19-MAY-1998  (first entry)
DE  Human partial mature membrane protein.
KW  Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW  brain; nervous tissue; cancer.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  Protein         1..612
FT                  /note= "partial mature protein"
PN  J10036395-A.
PD  10-FEB-1998.
PF  24-JUL-1996; 194467.
PR  24-JUL-1996; JP-194467.
PA  (ASAH ) ASAHI KASEI KOGYO KK.
DR  WPI; 98-174912/16.
PT  New human membrane protein - specifically expressed in brain and
PT  nervous tissue; used in diagnosis of diseases specific to these
PT  tissues and cancer
PS  Claim 1; Column 18-19; 26pp; Japanese.
CC  W39256 represents the partial mature amino acid sequence of a novel
CC  membrane protein which contains epidermal growth factor (EGF) motifs.
CC  The new membrane protein is expressed specifically in brain and nervous
CC  tissue. The protein and DNA can be used in the diagnosis of brain and
CC  nerve system specific diseases and cancer.
SQ  Sequence  612 AA;

  Query Match          13.7%;  Score 120;  DB 28;  Length 612;
  Best Local Similarity  48.6%;  Pred. No. 5.31e-02;
  Matches   18;  Conservative    7;  Mismatches   8;  Indels   4;  Gaps   4;

Db    451 cahgtcr-sv-gtsykclcdpgyhglyceeeynecls 485
          | || |: |  | :| |||:||: |  ||:| | |::
Qy      4 CHHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMG 38


RESULT  12
ID  W39257 standard; protein; 737 AA.
AC  W39257;
DT  19-MAY-1998  (first entry)
DE  Human membrane protein.
KW  Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW  brain; nervous tissue; cancer; disease.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  Peptide         1..26
FT                  /label= signal
FT  Protein         27..737
FT                  /label= membrane_protein
PN  J10036395-A.
PD  10-FEB-1998.
PF  24-JUL-1996; 194467.
PR  24-JUL-1996; JP-194467.
PA  (ASAH ) ASAHI KASEI KOGYO KK.
DR  WPI; 98-174912/16.
DR  N-PSDB; V09641.
```

```
PT   New human membrane protein - specifically expressed in brain and
PT   nervous tissue; used in diagnosis of diseases specific to these
PT   tissues and cancer
PS   Claim 2; Pages 19-21; 26pp; Japanese.
CC   W39257 represents the amino acid sequence of a novel membrane protein
CC   which contains epidermal growth factor (EGF) motifs. The new membrane
CC   protein is expressed specifically in brain and nervous tissue. The
CC   protein and DNA can be used in the diagnosis of brain and nerve system
CC   specific diseases and cancer.
SQ   Sequence   737 AA;

  Query Match           13.7%;  Score 120;  DB 28;  Length 737;
  Best Local Similarity  48.6%;  Pred. No. 5.31e-02;
  Matches   18;  Conservative    7;  Mismatches   8;  Indels   4;  Gaps   4;

Db    477 cahgtcr-sv-gtsykclcdpgyhglyceeeynecls 511
          | || |: |  | :| |||:||: |  ||:| | |::
Qy      4 CHHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMG 38


RESULT   13
     R64230 standard; Protein; 121 AA.
     R64230;
DT   24-AUG-1995  (first entry)
DE   HCT-15 contg. hTG-700 tumour cell proliferation inhibiting factor.
KW   HCT-15; tumour cell proliferation inhibiting factor; hTG-700;
KW   leukaemia; renal cancer; cervical cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   protein         15..60
FT                   /label= hTG-700
PN   WO9429340-A.
PD   22-DEC-1994.
PF   02-JUN-1994; J00895.
PR   04-JUN-1993; JP-134854.
PR   04-JUN-1993; JP-134855.
PA   (TAIS ) TAISHO PHARM CO LTD.
PI   Hanada K,  Komurasaki T,  Nakazawa K,  Takahashi M;
PI   Toyoda H,  Uchida D,  Udaka S;
DR   WPI; 95-036401/05.
DR   N-PSDB; Q80358.
PT   Human tumour cell proliferation inhibiting factor  - for the
PT   treatment of leukaemia, renal cancer, cervical cancer etc.
PS   Claim 1; Fig 4; 62pp; Japanese.
CC   Q80358 encodes R64230 HCT-15, which contains the human tumour
CC   cell proliferation inhibiting factor hTG-700 (R64229). The factor
CC   may be useful in remedies for leukaemia, renal cancer and cervical
CC   cancer.
SQ   Sequence   121 AA;

     uery Match           13.5%;  Score 118;  DB 12;  Length 121;
     est Local Similarity  43.3%;  Pred. No. 7.73e-02;
  Matches   13;  Conservative    8;  Mismatches   8;  Indels   1;  Gaps   1;

Db     28 clhgqciylvdmsqnycrcevgytgvrceh 57
          | ||||  : | :: || |: |::| :||:
Qy      4 CHHGQC-HISDRGEPYCLCQPGFSGHHCEQ 32


RESULT   14
ID   W35957 standard; Protein; 381 AA.
AC   W35957;
DT   05-MAR-1998  (first entry)
DE   Human monocyte mature differentiation factor.
KW   Human; monocyte; mature; differentiation factor; MMDF; macrophage;
KW   cancer; immune activator; tissue culture; infectious disease.
OS   Homo sapiens.
PN   J09234079-A.
PD   09-SEP-1997.
PF   04-MAR-1996; 075236.
PR   04-MAR-1996; JP-075236.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 97-497320/46.
DR   N-PSDB; T97142.
PT   A monocyte mature differentiation factor - useful for the long term
PT   tissue culture of macrophage(s)
PS   Claim 9; Page 12-13; 22pp; Japanese.
CC   The present sequence represents a monocyte mature differentiation
CC   factor (MMDF) which maintains the life of macrophages for long periods
CC   in liquid culture. MMDF can be used as an anti-cancer agent, an immune
CC   activator and to treat infectious diseases.
SQ   Sequence   381 AA;

  Query Match           13.1%;  Score 114;  DB 26;  Length 381;
  Best Local Similarity  32.3%;  Pred. No. 1.63e-01;
  Matches   20;  Conservative   10;  Mismatches  29;  Indels   3;  Gaps   3;

Db    300 yagclsvkkyrpkyc-gscvdgrcctpqltrtvkmrfrcedgetfsknvmmiqsckcnyn 358
          ||:| :  |   : | |:|    || |  ::  |  |:| || :|  :|     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    359 cp 360
          |:
Qy    109 CS 110


RESULT   15
ID   W35730 standard; Protein; 381 AA.
AC   W35730;
DT   27-MAR-1998  (first entry)
DE   Human cysteine rich protein 61 (Cyr61).
KW   Cysteine rich protein 61; Cyr61; human;
KW   extracellular matrix signalling molecule; cell adhesion;
KW   cell migration; cell proliferation; angiogenesis; chondrogenesis;
KW   oncogenesis; haematostasis; wound healing; organ regeneration.
OS   Homo sapiens.
PN   WO9733995-A2.
PD   18-SEP-1997.
PF   14-MAR-1997; U04193.
PR   15-MAR-1996; US-013958.
PA   (MUNI-) MUNIN CORP.
PI   Lau LF;
DR   WPI; 97-470875/43.
DR   N-PSDB; T94699.
PT   Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT   modulate e.g. haematostasis, induce wound healing, promote organ
PT   regeneration etc
PS   Claim 2; Page 112-113; 133pp; English.
CC   This protein sequence comprises human cysteine rich protein 61
CC   (Cyr61), an extracellular matrix signalling molecule.  Its amino
CC   acid sequence was deduced from a human placental cDNA clone (see
CC   T94699).  Cyr61 polypeptides can be expressed in transformed or
CC   transfected host cells.  Cyr61 can be used to modulate
CC   haematostasis, induce wound healing in a tissue, promote organ
CC   regeneration, improve tissue grafting or promote bone or prothesis
CC   implantation (claimed).  It can also be used to screen for a
CC   modulator of angiogenesis, chondrogenesis, oncogenesis, cell
CC   adhesion, cell migration, cell proliferation, expand a population
CC   of undifferentiated haematopoietic stem cells in culture and to
CC   screen for a mitogen (claimed).  Ex vivo methods for using
CC   mammalian extracellular matrix signalling molecules to prepare
CC   blood products are also provided.
SQ   Sequence   381 AA;

  Query Match           13.1%;  Score 114;  DB 26;  Length 381;
  Best Local Similarity  32.3%;  Pred. No. 1.63e-01;
  Matches   20;  Conservative   10;  Mismatches  29;  Indels   3;  Gaps   3;

Db    300 yagclsvkkyrpkyc-gscvdgrcctpqltrtvkmrfrcedgetfsknvmmiqsckcnyn 358
          ||:| :  |   : | |:|    || |  ::  |  |:| || :|  :|     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    359 cp 360
          |:
Qy    109 CS 110
```

Search completed: Fri May 28 09:24:32 1999
Job time : 51 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:24:51 1999; MasPar time 7.38 Seconds
597.165 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-11
Description: (1-110) from US09191647.pep
Perfect Score: 873
Sequence: 1 AFKCHHGQCHISDRGEPYCL.....GSSFVEEVERHLECGCRACS 110

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 37.070; Variance 65.147; scale 0.569

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	181	20.7	1469	2	B36665	slit protein 2 precur	2.45e-18
2	167	19.1	1480	2	A36665	slit protein 1 precur	1.43e-15
3	133	15.2	2531	2	A46019	gene Notch-1 protein	3.98e-09
4	131	15.0	2321	2	S78549	notch3 protein - huma	9.21e-09
5	131	15.0	2437	2	S42612	transmembrane protein	9.21e-09
6	127	14.5	375	2	A41428	CEF-10 protein precur	4.86e-08
7	126	14.4	379	2	A35669	gene CYR61 protein pr	7.35e-08
8	126	14.4	5147	1	IJFFTM	cadherin-related tumo	7.35e-08
9	125	14.3	293	2	B26637	neurogenic repetitive	1.11e-07
10	125	14.3	2139	2	A35672	crumbs protein - frui	1.11e-07
11	125	14.3	2531	2	S18188	notch protein homolog	1.11e-07
12	125	14.3	2555	2	A40043	notch protein homolog	1.11e-07
13	122	14.0	2703	2	A24420	notch protein - fruit	3.79e-07
14	121	13.9	2524	2	A35844	Xotch protein - Afric	5.69e-07
15	118	13.5	4391	2	A38096	perlecan precursor -	1.91e-06
16	117	13.4	530	2	A31640	epidermal growth fact	2.86e-06
17	116	13.3	861	2	A48825	Notch homolog Motch p	4.26e-06
18	115	13.2	1203	2	A49175	Motch B protein - mou	6.35e-06
19	115	13.2	1429	2	S06434	homeotic protein lin-	6.35e-06
20	115	13.2	2471	2	A49128	cell-fate determining	6.35e-06
21	114	13.1	728	2	I50719	C-Delta-1 - chicken	9.45e-06
22	114	13.1	2318	2	S45306	notch 3 protein - mou	9.45e-06
23	113	12.9	177	2	A37408	betacellulin precurso	1.40e-05
24	110	12.6	612	2	S23174	endothelial leukocyte	4.56e-05
25	110	12.6	618	2	B42755	E-selectin precursor	4.56e-05
26	110	12.6	1955	1	AGCH	agrin precursor - chi	4.56e-05
27	109	12.5	46	2	JT0747	epiregulin - rat	6.73e-05
28	109	12.5	162	2	S68401	epiregulin precursor	6.73e-05
29	109	12.5	401	2	S65138	glycoprotein antigen	6.73e-05
30	109	12.5	427	2	S74211	PAS-6/7 protein precu	6.73e-05
31	109	12.5	2871	2	A55624	fibrillin-1 precursor	6.73e-05
32	109	12.5	3002	2	A47221	fibrillin 1 precursor	6.73e-05
33	108	12.4	259	2	S48713	fetal antigen 1 - hum	9.92e-05
34	108	12.4	260	2	A44549	fetal antigen 1 homeo	9.92e-05
35	108	12.4	383	2	S53716	homeotic protein dlk	9.92e-05
36	108	12.4	383	2	B45484	delta-like dlk homeot	9.92e-05
37	108	12.4	638	2	S08042	proteoglycan core pro	9.92e-05
38	107	12.3	385	2	S53718	homeotic protein dlk	1.46e-04
39	107	12.3	385	2	A54785	preadipocyte factor 1	1.46e-04
40	107	12.3	1295	2	A32901	glp1 protein precurso	1.46e-04
41	106	12.1	159	2	I57497	transforming growth f	2.15e-04
42	106	12.1	160	1	WFHU1	transforming growth f	2.15e-04
43	106	12.1	208	2	A38432	heparin-binding EGF-1	2.15e-04
44	106	12.1	208	2	A41914	diptheria toxin recep	2.15e-04
45	106	12.1	706	2	H71707	hypothetical protein	2.15e-04

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
  371-394         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494         #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537         #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695         #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733         #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  846-890         #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061       #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           20.7%;  Score 181;  DB 2;  Length 1469;
  Best Local Similarity  31.3%;  Pred. No. 2.45e-18;
  Matches   35;  Conservative   22;  Mismatches  46;  Indels   9;  Gaps   9;

Db   1361 KCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQAASTCRKEQVREYYT-ENDCRSRQPL- 1418
          ||::| :|   |:  : | | |  |  |: |:|  : |  | |||    ::|  |
Qy      3 KCHHG-QC-HISDRGEPY-CLCQPGFSGHHCEQENP-CMGEIVREAIRRQKDYASCATAS 58

Db   1419 KYA-K-CVGGCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGCTKKC 1468
          | : : | ||||  || :   :|||    |::   :: :::    ||| : |
Qy     59 KVPIMECRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGC-RAC 109


RESULT    2
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
  66-91         #domain proteoglycan amino-terminal homology #label
                  PAH1\
  101-124       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
  125-148       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
  149-172       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
  173-196       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
  197-220       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
  228-272       #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
  288-313       #domain proteoglycan amino-terminal homology #label
                  PAH2\
  323-346       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
  347-370       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
  371-394       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
  395-418       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
  419-442       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
  450-494       #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
  512-537       #domain proteoglycan amino-terminal homology #label
                  PAH3\
  547-571       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
  572-595       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
  596-619       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
  620-643       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695       #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733       #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  791-814       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
  815-838       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
  846-890       #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061     #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           19.1%;  Score 167;  DB 2;  Length 1480;
  Best Local Similarity  29.4%;  Pred. No. 1.43e-15;
  Matches   35;  Conservative   25;  Mismatches  47;  Indels  12;  Gaps  12;

Db   1361 KCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQGEGSTEPPTVTAASTCRKEQVREYYTE 1420
          ||::| :|   |:  : | | |  |  |: |:| | :     |  |   |: | ::| :
Qy      3 KCHHG-QC-HISDRGEPY-CLCQPGFSGHHCEQ-ENPCMGEIVREA-I-RR-Q-KDY-AS 53

Db   1421 NDCRSRQPLKYAKCVGGCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGCTKKC 1479
               |: |::   | ||||  || :   :|||    |::   :: :::    ||| : |
Qy     54 CATASKVPIM-E-CRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGC-RAC 109


RESULT    3
ENTRY           A46019     #type complete
TITLE           gene Notch-1 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
```

```
DATE            22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A46019
REFERENCE       A46019
  #authors      del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins,
                  N.A.; Copeland, N.G.; Gridley, T.
  #journal      Genomics (1993) 15:259-264
  #title        Cloning, analysis, and chromosomal localization of Notch-1, a
                  mouse homolog of Drosophila Notch.
  #cross-references MUID:93194170
  #accession    A46019
    ##status       preliminary; not compared with conceptual translation
    ##molecule_type nucleic acid
    ##residues     1-2531 ##label DEL
    ##cross-references GB:Z11886; GB:S47228; NID:g288502; PID:g288503
    ##note         sequence extracted from NCBI backbone (NCBIP:127318)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
  757-788           #domain EGF homology #label EGF\
  1917-1948         #domain ankyrin repeat homology #label AN1\
  1949-1981         #domain ankyrin repeat homology #label AN2\
  1983-2015         #domain ankyrin repeat homology #label AN3\
  2016-2048         #domain ankyrin repeat homology #label AN4\
  2049-2081         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2531  #molecular-weight 271312  #checksum 6611

 Query Match          15.2%;  Score 133;  DB 2;  Length 2531;
 Best Local Similarity  25.4%;  Pred. No. 3.98e-09;
 Matches   29;  Conservative   29;  Mismatches  47;  Indels   9;  Gaps   9;

Db    575 CHYGSCK-DGVATFTCLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNSYLCLCLKGT 633
          || | |     :   ||||||::||||| : | | ::  | :   : |:: |   |  ::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGEIVREA-I-R-RQKDYAS-CATAS 58

Db    634 TGPNCEINLDDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            | |     |::  |::     :   ::|:  :::   : : :: ||:  :|
Qy     59 KVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT    4
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
  #authors      Joutel, A.; Tournier-Lasserve, E.
  #submission   submitted to the EMBL Data Library, April 1997
  #accession    S78549
    ##molecule_type mRNA
    ##residues     1-2321 ##label JOU1
    ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE       S71825
  #authors      Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
  #journal      Nature (1996) 383:707-710
  #title        Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
  #cross-references MUID:97032728
  #accession    S71825
    ##status       nucleic acid sequence not shown
    ##molecule_type DNA
    ##residues     67-113;138-194;268-333,'G',335-346;536-613;716-765;
                    1240-1279;1815-1888 ##label JOU2
    ##cross-references EMBL:U97669
GENETICS
  #gene         notch3
  #map_position 19p13.1
FUNCTION
  #description  may be involved in pathogenesis of CADASIL, causing a type of
                  stroke and dementia
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        tandem repeat; transmembrane protein
FEATURE
  318-349           #domain EGF homology #label EGF\
  1838-1870         #domain ankyrin repeat homology #label AN1\
  1871-1903         #domain ankyrin repeat homology #label AN2\
  1905-1937         #domain ankyrin repeat homology #label AN3\
  1938-1970         #domain ankyrin repeat homology #label AN4\
  1971-2003         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2321  #molecular-weight 243657  #checksum 3337

 Query Match          15.0%;  Score 131;  DB 2;  Length 2321;
 Best Local Similarity  42.1%;  Pred. No. 9.21e-09;
 Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db    1055 CQAGGQC-VDEDSSHYCVCPEGRTGSHCEQEVDPCLAQ 1091
           |: | || : : :  ||:|  | :| ||||| :||:::
Qy       4 CHHG-QCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT    5
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
  #authors      Bierkamp, C.; Campos-Ortega, J.A.
  #journal      Mech. Dev. (1993) 43:87-100
  #title        A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
  #cross-references MUID:94128602
  #accession    S42612
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-2437 ##label BIE
    ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
  1915-1947         #domain ankyrin repeat homology #label AN1\
  1948-1980         #domain ankyrin repeat homology #label AN2\
  1982-2014         #domain ankyrin repeat homology #label AN3\
  2015-2047         #domain ankyrin repeat homology #label AN4\
  2048-2080         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

 Query Match          15.0%;  Score 131;  DB 2;  Length 2437;
 Best Local Similarity  41.0%;  Pred. No. 9.21e-09;
 Matches   16;  Conservative   12;  Mismatches   9;  Indels   2;  Gaps   2;

Db    1352 SLRCRNGATCVSGHLSPRCLCAPGFSGHECQTRMDSPCL 1390
           :::|::|   :| :  | ||| ||||||  |: : ::||:
Qy       1 AFKCHHGQCHISDRGEPYCLCQPGFSGHHCE-Q-ENPCM 37


RESULT    6
ENTRY           A41428     #type complete
TITLE           CEF-10 protein precursor - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
DATE            03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
                  10-Sep-1997
ACCESSIONS      A41428
REFERENCE       A41428
  #authors      Simmons, D.L.; Levy, D.B.; Yannoni, Y.; Erikson, R.L.
  #journal      Proc. Natl. Acad. Sci. U.S.A. (1989) 86:1178-1182
  #title        Identification of a phorbol ester-repressible v-src-inducible
```

```
                gene.
   #cross-references MUID:89145206
   #accession    A41428
     ##status       preliminary
     ##molecule_type mRNA
     ##residues     1-375 ##label SIM
     ##cross-references GB:J04496; NID:g211435; PID:g211436
SUMMARY          #length 375  #molecular-weight 40651  #checksum 1417

  Query Match           14.5%;  Score 127;  DB 2;  Length 375;
  Best Local Similarity  32.3%;  Pred. No. 4.86e-08;
  Matches    20;  Conservative   11;  Mismatches  28;  Indels   3;  Gaps   3;

Db    295 YAGCSSVKKYRPKYC-GSCVDGRCCTPQQTRTVKIRFRCDDGETFTKSVMMIQSCRCNYN 353
        ||:|:: |   : | |:|    || | ::: |  |:| || :|   |     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    354 CP 355
          |:
Qy    109 CS 110


RESULT    7
ENTRY            A35669     #type complete
TITLE            gene CYR61 protein precursor - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                   16-Dec-1998
ACCESSIONS       A35669; I48319; S16446
REFERENCE        A35669
   #authors      O'Brien, T.P.; Yang, G.P.; Sanders, L.; Lau, L.F.
   #journal      Mol. Cell. Biol. (1990) 10:3569-3577
   #title        Expression of cyr61, a growth factor-inducible
                   immediate-early gene.
   #cross-references MUID:90287146
   #accession    A35669
     ##status       preliminary
     ##molecule_type mRNA
     ##residues     1-379 ##label OAB
     ##cross-references GB:M32490; NID:g192909; PID:g309206
     ##note         the authors translated the codon GAT for residue 337 as
                     Gln
REFERENCE        I48319
   #authors      Latinkic, B.V.; O'Brien, T.P.; Lau, L.F.
   #journal      Nucleic Acids Res. (1991) 19:3261-3267
   #title        Promoter function and structure of the growth
                   factor-inducible immediate early gene cyr61.
   #cross-references MUID:91288203
   #accession    I48319
     ##status       translated from GB/EMBL/DDBJ
     ##molecule_type DNA
     ##residues     1-379 ##label RES
     ##cross-references EMBL:X56790; NID:g50632; PID:g50633
     ##note         the authors did not translate the codon for residue 108
     ##note         the authors translated the codon GAT for residue 337 as
                     Gln
GENETICS
   #gene         CYR61
   #introns      21/3; 93/1; 208/1; 279/3
CLASSIFICATION   #superfamily von Willebrand factor type C repeat homology
FEATURE
   99-166            #domain von Willebrand factor type C repeat homology
                       #label VWC
SUMMARY          #length 379  #molecular-weight 41709  #checksum 3726

  Query Match           14.4%;  Score 126;  DB 2;  Length 379;
  Best Local Similarity  32.3%;  Pred. No. 7.35e-08;
  Matches    20;  Conservative   12;  Mismatches  27;  Indels   3;  Gaps   3;

Db    298 YAGCSSVKKYRPKYC-GSCVDGRCCTPLQTRTVKMRFRCEDGEMFSKNVMMIQSCKCNYN 356
        ||:|:: |   : | |:|    || |:::: |  |:| ||  |  :|    | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    357 CP 358
          |:
Qy    109 CS 110


RESULT    8
ENTRY            IJFFTM     #type complete
TITLE            cadherin-related tumor suppressor precursor - fruit fly
                   (Drosophila melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                   16-Feb-1997
ACCESSIONS       A41087; B41087
REFERENCE        A41087
   #authors      Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.;
                   Bryant, P.J.; Goodman, C.S.
   #journal      Cell (1991) 67:853-868
   #title        The fat tumor suppressor gene in Drosophila encodes a novel
                   member of the cadherin gene superfamily.
   #cross-references MUID:92069752
   #accession    A41087
     ##molecule_type mRNA
     ##residues     143-485;1279-5147 ##label MAH
     ##cross-references GB:M80537
   #accession    B41087
     ##molecule_type DNA
     ##residues     1-142;487-1278 ##label MA2
     ##cross-references GB:M80537
     ##note         1229-Gly and 1233-Ser were also found
GENETICS
   #gene         fat
     ##cross-references FlyBase:FBgn0001075
CLASSIFICATION   #superfamily cadherin-related tumor suppressor; cadherin
                   repeat homology; EGF homology
KEYWORDS         calcium binding; cell adhesion; duplication; transmembrane
                   protein
FEATURE
   1-35              #domain signal sequence #status predicted #label SIG\
   36-5147           #product cadherin-related tumor suppressor #status
                       predicted #label MAT\
   36-4583           #domain extracellular #status predicted #label EXT\
   51-156            #domain cadherin repeat homology #label CR1\
   159-270           #domain cadherin repeat homology #label CR2\
   271-382           #domain cadherin repeat homology #label CR3\
   390-494           #domain cadherin repeat homology #label CR4\
   497-599           #domain cadherin repeat homology #label CR5\
   602-708           #domain cadherin repeat homology #label CR6\
   718-822           #domain cadherin repeat homology #label CR7\
   831-942           #domain cadherin repeat homology #label CR8\
   948-1049          #domain cadherin repeat homology #label CR9\
   1052-1153         #domain cadherin repeat homology #label C10\
   1156-1278         #domain cadherin repeat homology #label C11\
   1281-1384         #domain cadherin repeat homology #label C12\
   1387-1489         #domain cadherin repeat homology #label C13\
   1492-1601         #domain cadherin repeat homology #label C14\
   1607-1713         #domain cadherin repeat homology #label C15\
   1717-1823         #domain cadherin repeat homology #label C16\
   1826-1922         #domain cadherin repeat homology #label C17\
   1925-2027         #domain cadherin repeat homology #label C18\
   2028-2167         #domain cadherin repeat homology #label C19\
   2169-2278         #domain cadherin repeat homology #label C20\
   2281-2384         #domain cadherin repeat homology #label C99\
   2387-2491         #domain cadherin repeat homology #label C21\
   2494-2596         #domain cadherin repeat homology #label C22\
   2599-2703         #domain cadherin repeat homology #label C23\
   2707-2810         #domain cadherin repeat homology #label C24\
   2813-2913         #domain cadherin repeat homology #label C25\
   2915-3013         #domain cadherin repeat homology #label C26\
   3014-3124         #domain cadherin repeat homology #label C27\
   3127-3229         #domain cadherin repeat homology #label C28\
   3232-3334         #domain cadherin repeat homology #label C29\
   3337-3439         #domain cadherin repeat homology #label C30\
   3442-3545         #domain cadherin repeat homology #label C31\
```

```
   3548-3651          #domain cadherin repeat homology #label C32\
   3654-3756          #domain cadherin repeat homology #label C33\
   3954-4010          #domain EGF homology #label EG1\
   4017-4048          #domain EGF homology #label EG2\
   4056-4089          #domain EGF homology #label EG3\
   4096-4127          #domain EGF homology #label EG4\
   4584-4609          #domain transmembrane #status predicted #label TMM\
   4610-5147          #domain intracellular #status predicted #label INT
SUMMARY          #length 5147  #molecular-weight 564895  #checksum 6994

  Query Match           14.4%;  Score 126;  DB 1;  Length 5147;
  Best Local Similarity  44.4%;  Pred. No. 7.35e-08;
  Matches   16;  Conservative    9;  Mismatches   8;  Indels   3;  Gaps   3;

Db   4061 CRNGGSCQRSPDGSSYFCLCRPGFRGNQCESVSDSC 4096
         |::|  |: |  | :| |||:||| |::||   ::|
Qy      4 CHHG-QCHISDRGEPY-CLCQPGFSGHHCEQ-ENPC 36


RESULT    9
ENTRY            B26637     #type fragment
TITLE            neurogenic repetitive locus 95F protein - fruit fly
                   (Drosophila melanogaster) (fragment)
ORGANISM         #formal_name Drosophila melanogaster
DATE             16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
                   14-Aug-1998
ACCESSIONS       B26637
REFERENCE        A91081
   #authors      Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigel,
                   D.; Vaessin, H.; Campos-Ortega, J.A.
   #journal      EMBO J. (1987) 6:761-766
   #title        EGF homologous sequences encoded in the genome of Drosophila
                   melanogaster, and their relation to neurogenic genes.
   #cross-references MUID:87218537
   #accession    B26637
      ##molecule_type mRNA
      ##residues      1-293 ##label KNU
      ##cross-references GB:X05144; NID:g7519; PID:g929536
GENETICS
   #gene         FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         transmembrane protein
FEATURE
   216-252           #domain EGF homology #label EGF
SUMMARY          #length 293  #checksum 3413

  Query Match           14.3%;  Score 125;  DB 2;  Length 293;
  Best Local Similarity  42.1%;  Pred. No. 1.11e-07;
  Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db    144 CLNNGTC-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 180
          | ::| | |:: :  :| ||||| |:|||: : |  :
Qy      4 C-HHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT   10
ENTRY            A35672     #type complete
TITLE            crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                   14-Aug-1998
ACCESSIONS       A35672
REFERENCE        A35672
   #authors      Tepass, U.; Theres, C.; Knust, E.
   #journal      Cell (1990) 61:787-799
   #title        crumbs encodes an EGF-like protein expressed on apical
                   membranes of Drosophila epithelial cells and required for
                   organization of epithelia.
   #cross-references MUID:90263104
   #accession    A35672
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-2139 ##label TEP
      ##cross-references GB:M33753
      ##note         the authors translated the codon GGC for residue 1928 as
                      Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene         FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         transmembrane protein
FEATURE
   691-722           #domain EGF homology #label EGF
SUMMARY          #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           14.3%;  Score 125;  DB 2;  Length 2139;
  Best Local Similarity  42.1%;  Pred. No. 1.11e-07;
  Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db   1806 CLNNGTC-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 1842
          | ::| | |:: :  :| ||||| |:|||: : |  :
Qy      4 C-HHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT   11
ENTRY            S18188     #type complete
TITLE            notch protein homolog - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
                   12-Feb-1999
ACCESSIONS       S18188
REFERENCE        S18188
   #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal      Development (1991) 113:199-205
   #title        A homolog of Drosophila Notch expressed during mammalian
                   development.
   #cross-references MUID:92111383
   #accession    S18188
      ##molecule_type mRNA
      ##residues      1-2531 ##label WEI
      ##cross-references EMBL:X57405; NID:g57634; PID:g57635
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1917-1949         #domain ankyrin repeat homology #label AN1\
   1950-1982         #domain ankyrin repeat homology #label AN2\
   1984-2016         #domain ankyrin repeat homology #label AN3\
   2017-2049         #domain ankyrin repeat homology #label AN4\
   2050-2082         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2531  #molecular-weight 270907  #checksum 2705

  Query Match           14.3%;  Score 125;  DB 2;  Length 2531;
  Best Local Similarity  25.4%;  Pred. No. 1.11e-07;
  Matches   30;  Conservative   31;  Mismatches  45;  Indels  12;  Gaps  11;

Db    570 CDPDPCHIGLCKDGVATFTCLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNYYLCLC 629
          |   |||:  : |   : ||||||::||||| : | | ::  | :   :::|  |   |
Qy      4 CHHGQCHISD-R-G-EPY-CLCQPGFSGHHCEQE-NPCMGEIVREA-IRRQKD--YAS-C 54

Db    630 LKGTTGPNCEINLDDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            ::  | |    |::  |::     :   ::|:  :::   : : :: ||:  :|
Qy     55 ATASKVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT   12
ENTRY            A40043     #type complete
TITLE            notch protein homolog TAN-1 precursor - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
                   14-Aug-1998
ACCESSIONS       A40043
REFERENCE        A40043
   #authors      Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds,
                   T.C.; Smith, S.D.; Sklar, J.
   #journal      Cell (1991) 66:649-661
```

```
   #title       TAN-1, the human homolog of the Drosophila Notch gene, is
                  broken by chromosomal translocations in T lymphoblastic
                  neoplasms.
   #cross-references MUID:91347367
   #accession   A40043
     ##status       preliminary; nucleic acid sequence not shown; not
                     compared with conceptual translation
     ##molecule_type mRNA
     ##residues     1-2555 ##label ELL
     ##cross-references GB:M73980
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   1149-1180         #domain EGF homology #label EGF\
   1927-1959         #domain ankyrin repeat homology #label AN1\
   1960-1992         #domain ankyrin repeat homology #label AN2\
   1994-2026         #domain ankyrin repeat homology #label AN3\
   2027-2059         #domain ankyrin repeat homology #label AN4\
   2060-2092         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2555  #molecular-weight 272337  #checksum 463

 Query Match           14.3%;  Score 125;  DB 2;  Length 2555;
 Best Local Similarity  23.7%;  Pred. No. 1.11e-07;
 Matches   27;  Conservative   32;  Mismatches  46;  Indels   9;  Gaps   8;

Db    574 CHYGSCK-DGVATFTCLCRPGYTGHHCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGT 632
          || | |     :   |||:||::||||| : | | ::   :| : :  :: |   |  ::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE--IVREAIR-RQKDYAS-CATAS 58

Db    633 TGPNCEINLDDCASSPCDSGTCLDKIDGYECACEPGYTGSMCNSNIDECAGNPC 686
            | |     |::: |::     :   ::|:  :::   : : :: ||:  :|
Qy     59 KVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT   13
ENTRY           A24420     #type complete
TITLE           notch protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  neurogenic repetitive locus protein
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
                  07-Aug-1998
ACCESSIONS      A24420; A24768; S09358; A05267
REFERENCE       A24420
   #authors     Kidd, S.; Kelley, M.R.; Young, M.W.
   #journal     Mol. Cell. Biol. (1986) 6:3094-3108
   #cross-references MUID:87064624
   #accession   A24420
     ##molecule_type DNA
     ##residues     1-2703 ##label KID
     ##cross-references GB:K03508; NID:g157991; PID:g157993
REFERENCE       A24768
   #authors     Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
   #journal     Cell (1985) 43:567-581
   #cross-references MUID:86079539
   #accession   A24768
     ##molecule_type mRNA
     ##residues     1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',
                     268-872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',
                     2258-2264,'V',2266-2406,'R',2408-2444,'L',2446-2703
                     ##label WHA1
     ##note         the authors translated the codon ATC for residue 49 as
                     Thr, ATT for residue 2044 as Arg, GTA for residue 2265
                     as Ala, CGC for residue 2407 as His, and CTT for
                     residue 2445 as Arg
REFERENCE       S09358
   #authors     Tautz, D.
   #journal     Nucleic Acids Res. (1989) 17:6463-6471
   #title       Hypervariability of simple sequences as a general source for
                  polymorphic DNA markers.
   #cross-references MUID:89385974
   #accession   S09358
     ##molecule_type DNA
     ##residues     2505-2551,'QQQQ',2552-2576,'E',2578-2604 ##label TAU
REFERENCE       A05267
   #authors     Wharton, K.A.; Yedvobnick, B.; Pinnerty, V.G.;
                  Artavanis-Tsakonas, S.
   #journal     Cell (1985) 40:55-62
   #title       opa: a novel family of transcribed repeats shared by the
                  Notch locus and other developmentally regulated loci in D.
                  melanogaster.
   #cross-references MUID:85099329
   #accession   A05267
     ##molecule_type DNA
     ##residues     2504-2576,'E',2578-2611 ##label WHA2
GENETICS
   #gene        notch; opa
     ##cross-references FlyBase:FBgn0004647
   #map_position 8.96-9.36
   #introns     53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
CLASSIFICATION  #superfamily notch protein; ankyrin repeat homology; EGF
                  homology
KEYWORDS        differentiation; tandem repeat; transmembrane protein
FEATURE
   27-43             #domain transmembrane #status predicted #label TMM1\
   568-599           #domain EGF homology #label EGF\
   1746-1762         #domain transmembrane #status predicted #label TMM2\
   1950-1982         #domain ankyrin repeat homology #label AN1\
   1983-2015         #domain ankyrin repeat homology #label AN2\
   1988-2004         #domain transmembrane #status predicted #label TMM3\
   2017-2049         #domain ankyrin repeat homology #label AN3\
   2050-2082         #domain ankyrin repeat homology #label AN4\
   2083-2115         #domain ankyrin repeat homology #label AN5\
   2538-2568         #region glutamine-rich\
   2538-2568         #domain neurogenic repetitive element #status predicted
                       #label OPA
SUMMARY         #length 2703  #molecular-weight 288876  #checksum 6404

  Query Match           14.0%;  Score 122;  DB 2;  Length 2703;
  Best Local Similarity  40.5%;  Pred. No. 3.79e-07;
  Matches   17;  Conservative    9;  Mismatches  11;  Indels   5;  Gaps   5;

Db    952 SFPCQNGGTC-LDGIGD-YSCLCVDGFDGKHCETDINECLSQ 991
          :| |::|  | :   |: | |||  || | ||| : | |:::
Qy      1 AFKCHHG-QCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGE 39


RESULT   14
ENTRY           A35844     #type complete
TITLE           Xotch protein - African clawed frog
ORGANISM        #formal_name Xenopus laevis #common_name African clawed frog
DATE            12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change
                  14-Aug-1998
ACCESSIONS      A35844
REFERENCE       A35844
   #authors     Coffman, C.; Harris, W.; Kintner, C.
   #journal     Science (1990) 249:1438-1441
   #title       Xotch, the Xenopus homolog of Drosophila notch.
   #cross-references MUID:90385285
   #accession   A35844
     ##status       preliminary; nucleic acid sequence not shown; not
                     compared with conceptual translation
     ##molecule_type mRNA
     ##residues     1-2524 ##label COF
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        transmembrane protein
FEATURE
   222-254           #domain EGF homology #label EGF\
   1924-1956         #domain ankyrin repeat homology #label AN1\
   1957-1989         #domain ankyrin repeat homology #label AN2\
   1991-2023         #domain ankyrin repeat homology #label AN3\
   2024-2056         #domain ankyrin repeat homology #label AN4\
   2057-2089         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2524  #molecular-weight 274931  #checksum 9441

  Query Match           13.9%;  Score 121;  DB 2;  Length 2524;
```

```
  Best Local Similarity  45.0%;  Pred. No. 5.69e-07;
  Matches   18;  Conservative   8;  Mismatches  9;  Indels  5;  Gaps  5;

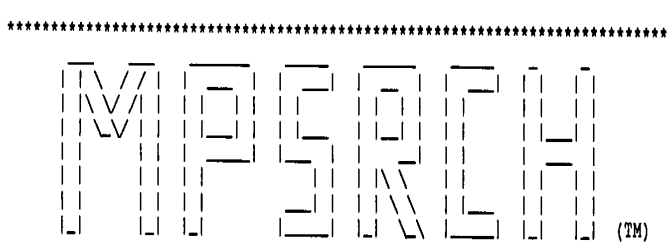
Db   1237 KCFNNGKC-I-DRVGGYNCICPPGFVGERCEGDVNECLSN 1274
          || ::| | | ||   | |:| ||| | :|| : | |:::
Qy      3 KC-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGE 39


RESULT  15
ENTRY           A38096     #type complete
TITLE           perlecan precursor - human
ALTERNATE_NAMES basement membrane heparan sulfate proteoglycan; heparan
                  sulfate proteoglycan 2
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
                  15-Jan-1999
ACCESSIONS      A38096; S19256; S77946; A41059; A40306; B33625; A33625;
                  A41736
REFERENCE       A38096
  #authors      Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo,
                  R.V.
  #journal      J. Biol. Chem. (1992) 267:8544-8557
  #title        Primary structure of the human heparan sulfate proteoglycan
                  from basement membrane (HSPG2/perlecan). A chimeric
                  molecule with multiple domains homologous to the low
                  density lipoprotein receptor, laminin, neural cell adhesion
                  molecules, and epidermal growth factor.
  #cross-references MUID:92235084
  #accession    A38096
    ##molecule_type mRNA
    ##residues     1-4391 ##label MUR
    ##cross-references GB:M85289; NID:g184426; PID:g184427
REFERENCE       A41736
  #authors      Kallunki, P.; Tryggvason, K.
  #journal      J. Cell Biol. (1992) 116:559-571
  #title        Human basement membrane heparan sulfate proteoglycan core
                  protein: a 467-kD protein containing multiple domains
                  resembling elements of the low density lipoprotein
                  receptor, laminin, neural cell adhesion molecules, and
                  epidermal growth factor.
  #cross-references MUID:92112994
  #accession    S19256
    ##molecule_type mRNA
    ##residues     1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',
                    503-792,'K',794-908,'R',910-1702,'RG',1705-1752,'R',
                    1754-2037,'I',2039-2049,'Q',2050-2051,2053-2092,'H',
                    2094-2626,'R',2628-2769,'Y',2771-2979,'H',2981-2994,
                    'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',
                    3428-3631,'Q',3633-3967,'S',3969-4003,'T',4005-4134,
                    'I',4136-4331,'I',4333-4391 ##label KAL
    ##cross-references EMBL:X62515
REFERENCE       S77946
  #authors      Tryggvason, K.
  #submission   submitted to the EMBL Data Library, October 1991
  #accession    S77946
    ##molecule_type mRNA
    ##residues     1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',
                    503-792,'K',794-908,'R',910-1702,'RG',1705-1752,'R',
                    1754-2037,'I',2039-2049,'Q',2050-2051,2053-2092,'H',
                    2094-2626,'R',2628-2769,'Y',2771-2979,'H',2981-2994,
                    'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',
                    3428-3631,'Q',3633-4003,'T',4005-4134,'I',4136-4331,
                    'I',4333-4391 ##label TRY
    ##cross-references EMBL:X62515; NID:g29469; PID:g29470
REFERENCE       A41059
  #authors      Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows,
                  T.B.; Tryggvason, K.
  #journal      Genomics (1991) 11:389-396
  #title        Cloning of human heparan sulfate proteoglycan core protein,
                  assignment of the gene (HSPG2) to 1p36.1->p35 and
                  identification of a BamHI restriction fragment length
                  polymorphism.
  #cross-references MUID:92120660
  #accession    A41059
    ##molecule_type mRNA
    ##residues     'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,
                    'L',1223-1397 ##label KA2
    ##cross-references GB:S76436; NID:g243370; PID:g243371
REFERENCE       A40306
  #authors      Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.;
                  McBride, O.W.; Yi, H.F.; Iozzo, R.V.
  #journal      Genomics (1991) 10:673-680
  #title        Heparan sulfate proteoglycan of human colon: partial
                  molecular cloning, cellular expression, and mapping of the
                  gene (HSPG2) to the short arm of human chromosome 1.
  #cross-references MUID:91365376
  #accession    A40306
    ##molecule_type mRNA
    ##residues     1018-1405,'G',1407-1409,'G',1411-1465 ##label DOD
    ##cross-references GB:M64283; NID:g184424; PID:g184425
REFERENCE       A33625
  #authors      Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson,
                  M.; Cassiman, J.J.; van den Berghe, H.; David, G.
  #journal      J. Cell Biol. (1989) 109:3199-3211
  #title        Matrix-associated heparan sulfate proteoglycan: core
                  protein-specific monoclonal antibodies decorate the
                  pericellular matrix of connective tissue cells and the
                  stromal side of basement membranes.
  #cross-references MUID:90078352
  #accession    B33625
    ##molecule_type protein
    ##residues     1379-1384,'X',1386-1388,'X',1390-1398 ##label HE2
  #accession    A33625
    ##molecule_type protein
    ##residues     2166-2171,'X',2173-2175,'X',2177-2185 ##label HE3
    ##note         peptide potentially matches four different regions of
                    sequence shown
GENETICS
  #gene         GDB:HSPG2
    ##cross-references GDB:126372; OMIM:142461
  #map_position 1p36.1-1p36.1
CLASSIFICATION  #superfamily LDL receptor ligand-binding repeat homology;
                  laminin G repeat homology; laminin-type EGF-like homology
KEYWORDS        chondroitin sulfate proteoglycan; glycoprotein; heparan
                  sulfate; transmembrane protein
FEATURE
    1-21            #domain signal sequence #status predicted #label SIG\
    22-4391         #product perlecan #status predicted #label MAT\
    22-193          #domain I #label DOM1\
    194-530         #domain II #label DOM2\
    199-234         #domain LDL receptor ligand-binding repeat homology
                      #label LDL1\
    285-319         #domain LDL receptor ligand-binding repeat homology
                      #label LDL2\
    325-359         #domain LDL receptor ligand-binding repeat homology
                      #label LDL3\
    368-403         #domain LDL receptor ligand-binding repeat homology
                      #label LDL4\
    531-1676        #domain III #label DOM3\
    1563-1610       #domain laminin-type EGF-like homology #label EG7\
    1677-3686       #domain IV #label DOM4\
    2007-2034       #domain transmembrane #status predicted #label TRM\
    3687-4391       #domain V #label DOM5\
    3953-4106       #domain laminin G repeat homology #label LG2\
    4149-4151       #region motor neuron attachment (L-R-E) motif\
    4299-4301       #region motor neuron attachment (L-R-E) motif\
    65,71,76        #binding_site heparan sulfate (Ser) (covalent) #status
                      predicted\
    89,554,1755,2121,
    3072,3105,3279,
    3780,3836,4068  #binding_site carbohydrate (Asn) (covalent) #status
                      predicted\
    2995,3933,4179  #binding_site chondroitin sulfate (Ser) (covalent)
                      #status predicted
SUMMARY         #length 4391  #molecular-weight 468819  #checksum 7166
```

Query Match 13.5%; Score 118; DB 2; Length 4391;
Best Local Similarity 45.2%; Pred. No. 1.91e-06;
Matches 14; Conservative 10; Mismatches 4; Indels 3; Gaps 3;

```
Db   3853 CQNGGQCHDSESS-SYVCVCPAGFTGSRCEH 3882
          |::| ||| |: : :| |:| :||:| :||:
Qy      4 CHHG-QCHISDRGEPY-CLCQPGFSGHHCEQ 32
```

Search completed: Fri May 28 09:25:09 1999
Job time : 18 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

ssrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:25:28 1999; MasPar time 5.14 Seconds
605.413 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-11
Description: (1-110) from US09191647.pep
Perfect Score: 873
Sequence: 1 AFKCHHGQCHISDRGEPYCL.....GSSFVEEVERHLECGCRACS 110

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 38.241; Variance 58.972; scale 0.648

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	167	19.1	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	7.49e-18
2	133	15.2	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	1.34e-10
3	131	15.0	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	3.44e-10
4	127	14.5	375	1	CE10_CHICK	CEF-10 PROTEIN PRECURS	2.23e-09
5	126	14.4	379	1	CYR6_MOUSE	CYR61 PROTEIN PRECURSO	3.54e-09
6	126	14.4	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	3.54e-09
7	125	14.3	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	5.62e-09
8	125	14.3	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	5.62e-09
9	125	14.3	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	5.62e-09
10	122	14.0	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	2.23e-08
11	121	13.9	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	3.52e-08
12	121	13.9	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	3.52e-08
13	118	13.5	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	1.37e-07
14	115	13.2	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	5.24e-07
15	114	13.1	381	1	CYR6_HUMAN	CYR61 PROTEIN PRECURSO	8.18e-07
16	114	13.1	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	8.18e-07
17	113	12.9	177	1	BTC_MOUSE	BETACELLULIN PRECURSOR	1.27e-06
18	112	12.8	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	1.98e-06
19	110	12.6	612	1	LEM2_MOUSE	E-SELECTIN PRECURSOR (	4.76e-06
20	110	12.6	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	4.76e-06
21	109	12.5	427	1	MFGM_BOVIN	MILK FAT GLOBULE-EGF F	7.35e-06
22	109	12.5	2871	1	FBN1_HUMAN	FIBRILLIN 1 PRECURSOR.	7.35e-06
23	109	12.5	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	7.35e-06
24	108	12.4	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	1.13e-05
25	108	12.4	2415	1	PGCA_HUMAN	AGGRECAN CORE PROTEIN	1.13e-05
26	107	12.3	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	1.75e-05
27	107	12.3	611	1	LEM2_CANFA	E-SELECTIN PRECURSOR (	1.75e-05
28	107	12.3	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	1.75e-05
29	106	12.1	84	1	HBGF_PIG	HEPARIN-BINDING EGF-LI	2.68e-05
30	106	12.1	121	1	TGFA_MACMU	TRANSFORMING GROWTH FA	2.68e-05
31	106	12.1	159	1	TGFA_RAT	TRANSFORMING GROWTH FA	2.68e-05
32	106	12.1	160	1	TGFA_HUMAN	TRANSFORMING GROWTH FA	2.68e-05
33	106	12.1	160	1	TGFA_PIG	TRANSFORMING GROWTH FA	2.68e-05
34	106	12.1	178	1	BTC_HUMAN	BETACELLULIN PRECURSOR	2.68e-05
35	106	12.1	208	1	HBGF_HUMAN	HEPARIN-BINDING EGF-LI	2.68e-05
36	106	12.1	208	1	HBGF_CERAE	HEPARIN-BINDING EGF-LI	2.68e-05
37	106	12.1	2871	1	FBN1_BOVIN	FIBRILLIN 1 PRECURSOR	2.68e-05
38	105	12.0	159	1	TGFA_MOUSE	TRANSFORMING GROWTH FA	4.11e-05
39	105	12.0	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	4.11e-05
40	105	12.0	4543	1	LRP1_CHICK	LOW-DENSITY LIPOPROTEI	4.11e-05
41	104	11.9	50	1	TGFA_RABIT	TRANSFORMING GROWTH FA	6.30e-05
42	103	11.8	133	1	TGFA_SHEEP	TRANSFORMING GROWTH FA	9.62e-05
43	103	11.8	140	1	GRFA_VACCV	GROWTH FACTOR.	9.62e-05
44	103	11.8	969	1	PAC4_HUMAN	SUBTILISIN-LIKE PROTEA	9.62e-05
45	103	11.8	2813	1	VWF_HUMAN	VON WILLEBRAND FACTOR	9.62e-05

ALIGNMENTS

```
RESULT   1
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE  1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          19.1%;  Score 167;  DB 1;  Length 1480;
  Best Local Similarity  29.4%;  Pred. No. 7.49e-18;
  Matches   35;  Conservative   25;  Mismatches  47;  Indels  12;  Gaps  12;

Db   1361 KCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQGEGSTEPPTVTAASTCRKEQVREYYTE 1420
           ||::| :|   |:  : | | |  |  |: |:| | :     |  |   |: | ::| :
Qy      3 KCHHG-QC-HISDRGEPY-CLCQPGFSGHHCEQ-ENPCMGEIVREA-I-RR-Q-KDY-AS 53

Db   1421 NDCRSRQPLKYAKCVGGCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGCTKKC 1479
             |: |::   | ||||  || :   :|||    |::  :: :::    ||| : |
Qy     54 CATASKVPIM-E-CRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGC-RAC 109


RESULT    2
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ----------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2075       ANK MOTIF 6.
FT   CARBOHYD   888    888       POTENTIAL.
FT   CARBOHYD   959    959       POTENTIAL.
FT   CARBOHYD  1179   1179       POTENTIAL.
FT   CARBOHYD  1241   1241       POTENTIAL.
FT   CARBOHYD  1489   1489       POTENTIAL.
FT   CARBOHYD  1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match          15.2%;  Score 133;  DB 1;  Length 2531;
  Best Local Similarity  25.4%;  Pred. No. 1.34e-10;
  Matches   29;  Conservative   29;  Mismatches  47;  Indels   9;  Gaps   9;

Db    575 CHYGSCK-DGVATFTCLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNSYLCLCLKGT 633
          || | |    :  ||||||::||||| : | | ::  | :  : |:: |  | ::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGEIVREA-I-R-RQKDYAS-CATAS 58

Db    634 TGPNCEINLDDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            | |    |:: |::    :  ::|: :::  : : :: ||: :|
Qy     59 KVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT   3
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21   1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725   1747       POTENTIAL.
FT   DOMAIN    1748   2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     57       EGF-LIKE 1.
FT   DOMAIN      58     98       EGF-LIKE 2.
FT   DOMAIN     101    138       EGF-LIKE 3.
FT   DOMAIN     139    175       EGF-LIKE 4.
FT   DOMAIN     177    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     827    865       EGF-LIKE 22.
FT   DOMAIN     867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     981   1017       EGF-LIKE 26.
FT   DOMAIN    1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN     1057   1093      EGF-LIKE 28.
FT   DOMAIN     1095   1141      EGF-LIKE 29.
FT   DOMAIN     1143   1179      EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1181   1217      EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1219   1263      EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1265   1303      EGF-LIKE 33.
FT   DOMAIN     1305   1344      EGF-LIKE 34.
FT   DOMAIN     1346   1382      EGF-LIKE 35.
FT   DOMAIN     1385   1423      EGF-LIKE 36.
FT   DOMAIN     1446   1561      3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1486      LIN/NOTCH 1.
FT   REPEAT     1487   1520      LIN/NOTCH 2.
FT   REPEAT     1521   1561      LIN/NOTCH 3.
FT   DOMAIN     1861   2074      6 X ANK MOTIF REPEATS.
FT   REPEAT     1861   1891      ANK MOTIF 1.
FT   REPEAT     1892   1940      ANK MOTIF 1.
FT   REPEAT     1941   1974      ANK MOTIF 1.
FT   REPEAT     1975   2007      ANK MOTIF 1.
FT   REPEAT     2008   2040      ANK MOTIF 1.
FT   REPEAT     2041   2074      ANK MOTIF 1.
FT   DOMAIN     2265   2276      POLY-GLN (OPA-REPEAT).
FT   DISULFID     25     35      BY SIMILARITY.
FT   DISULFID     29     45      BY SIMILARITY.
FT   DISULFID     47     56      BY SIMILARITY.
FT   DISULFID     62     73      BY SIMILARITY.
FT   DISULFID     67     86      BY SIMILARITY.
FT   DISULFID     88     97      BY SIMILARITY.
FT   DISULFID    105    116      BY SIMILARITY.
FT   DISULFID    110    126      BY SIMILARITY.
FT   DISULFID    128    137      BY SIMILARITY.
FT   DISULFID    143    154      BY SIMILARITY.
FT   DISULFID    148    163      BY SIMILARITY.
FT   DISULFID    165    174      BY SIMILARITY.
FT   DISULFID    181    194      BY SIMILARITY.
FT   DISULFID    188    203      BY SIMILARITY.
FT   DISULFID    205    214      BY SIMILARITY.
FT   DISULFID    221    232      BY SIMILARITY.
FT   DISULFID    226    242      BY SIMILARITY.
FT   DISULFID    244    253      BY SIMILARITY.
FT   DISULFID    260    271      BY SIMILARITY.
FT   DISULFID    265    280      BY SIMILARITY.
FT   DISULFID    282    291      BY SIMILARITY.
FT   DISULFID    298    311      BY SIMILARITY.
FT   DISULFID    305    320      BY SIMILARITY.
FT   DISULFID    322    331      BY SIMILARITY.
FT   DISULFID    338    349      BY SIMILARITY.
FT   DISULFID    343    358      BY SIMILARITY.
FT   DISULFID    360    369      BY SIMILARITY.
FT   DISULFID    375    386      BY SIMILARITY.
FT   DISULFID    380    397      BY SIMILARITY.
FT   DISULFID    399    408      BY SIMILARITY.
FT   DISULFID    415    428      BY SIMILARITY.
FT   DISULFID    422    437      BY SIMILARITY.
FT   DISULFID    439    448      BY SIMILARITY.
FT   DISULFID    455    466      BY SIMILARITY.
FT   DISULFID    460    475      BY SIMILARITY.
FT   DISULFID    477    486      BY SIMILARITY.
FT   DISULFID    493    503      BY SIMILARITY.
FT   DISULFID    498    512      BY SIMILARITY.
FT   DISULFID    514    523      BY SIMILARITY.
FT   DISULFID    530    541      BY SIMILARITY.
FT   DISULFID    535    550      BY SIMILARITY.
FT   DISULFID    552    561      BY SIMILARITY.
FT   DISULFID    568    578      BY SIMILARITY.
FT   DISULFID    573    587      BY SIMILARITY.
FT   DISULFID    589    598      BY SIMILARITY.
FT   DISULFID    605    616      BY SIMILARITY.
FT   DISULFID    610    625      BY SIMILARITY.
FT   DISULFID    627    636      BY SIMILARITY.
FT   DISULFID    643    653      BY SIMILARITY.
FT   DISULFID    648    662      BY SIMILARITY.
FT   DISULFID    664    673      BY SIMILARITY.
FT   DISULFID    680    691      BY SIMILARITY.
FT   DISULFID    685    700      BY SIMILARITY.
FT   DISULFID    702    711      BY SIMILARITY.
FT   DISULFID    718    728      BY SIMILARITY.
FT   DISULFID    723    737      BY SIMILARITY.
FT   DISULFID    739    748      BY SIMILARITY.
FT   DISULFID    755    766      BY SIMILARITY.
FT   DISULFID    760    775      BY SIMILARITY.
FT   DISULFID    777    786      BY SIMILARITY.
FT   DISULFID    793    804      BY SIMILARITY.
FT   DISULFID    798    813      BY SIMILARITY.
FT   DISULFID    815    824      BY SIMILARITY.
FT   DISULFID    831    842      BY SIMILARITY.
FT   DISULFID    836    853      BY SIMILARITY.
FT   DISULFID    855    864      BY SIMILARITY.
FT   DISULFID    871    882      BY SIMILARITY.
FT   DISULFID    876    891      BY SIMILARITY.
FT   DISULFID    893    902      BY SIMILARITY.
FT   DISULFID    909    920      BY SIMILARITY.
FT   DISULFID    914    929      BY SIMILARITY.
FT   DISULFID    931    940      BY SIMILARITY.
FT   DISULFID    947    958      BY SIMILARITY.
FT   DISULFID    952    967      BY SIMILARITY.
FT   DISULFID    969    978      BY SIMILARITY.
FT   DISULFID   1023   1034      BY SIMILARITY.
FT   DISULFID   1028   1043      BY SIMILARITY.
FT   DISULFID   1045   1054      BY SIMILARITY.
FT   DISULFID   1061   1072      BY SIMILARITY.
FT   DISULFID   1066   1081      BY SIMILARITY.
FT   DISULFID   1083   1092      BY SIMILARITY.
FT   DISULFID   1099   1120      BY SIMILARITY.
FT   DISULFID   1114   1129      BY SIMILARITY.
FT   DISULFID   1131   1140      BY SIMILARITY.
FT   DISULFID   1147   1158      BY SIMILARITY.
FT   DISULFID   1152   1167      BY SIMILARITY.
FT   DISULFID   1169   1178      BY SIMILARITY.
FT   DISULFID   1185   1196      BY SIMILARITY.
FT   DISULFID   1190   1205      BY SIMILARITY.
FT   DISULFID   1207   1216      BY SIMILARITY.
FT   DISULFID   1223   1242      BY SIMILARITY.
FT   DISULFID   1236   1251      BY SIMILARITY.
FT   DISULFID   1253   1262      BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          15.0%;  Score 131;  DB 1;  Length 2437;
  Best Local Similarity  41.0%;  Pred. No. 3.44e-10;
  Matches   16;  Conservative   12;  Mismatches   9;  Indels   2;  Gaps   2;

Db   1352 SLRCRNGATCVSGHLSPRCLCAPGFSGHECQTRMDSPCL 1390
          :::|::|   :| :  | ||| ||||| |: : ::||:
Qy      1 AFKCHHGQCHISDRGEPYCLCQPGFSGHHCE-Q-ENPCM 37


RESULT    4
ID   CE10_CHICK     STANDARD;      PRT;   375 AA.
AC   P19336;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CEF-10 PROTEIN PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89145206.
RA   SIMMONS D.L., LEVY D.B., YANNONI Y., ERIKSON R.L.;
RT   "Identification of a phorbol ester-repressible v-src-inducible gene.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:1178-1182(1989).
CC   -!- FUNCTION: PROBABLE SECRETED REGULATORY PROTEIN.
CC   -!- INDUCTION: BY V-SRC.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
```

```
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04496; G211436; -.
DR   PIR; A41428; A41428.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1    22
FT   CHAIN       23   375       CEF-10 PROTEIN.
FT   DOMAIN      98   164       VWFC.
FT   DOMAIN     281   355       CTCK.
FT   DISULFID   281   318       BY SIMILARITY.
FT   DISULFID   298   332       BY SIMILARITY.
FT   DISULFID   309   348       BY SIMILARITY.
FT   DISULFID   312   350       BY SIMILARITY.
FT   DISULFID   317   354       BY SIMILARITY.
SQ   SEQUENCE   375 AA;  40651 MW;  68B4BC92 CRC32;

  Query Match          14.5%;  Score 127;  DB 1;  Length 375;
  Best Local Similarity  32.3%;  Pred. No. 2.23e-09;
  Matches   20;  Conservative   11;  Mismatches  28;  Indels   3;  Gaps   3;

Db    295 YAGCSSVKKYRPKYC-GSCVDGRCCTPQQTRTVKIRFRCDDGETFTKSVMMIQSCRCNYN 353
          ||:|:: |   : | |:|    || | ::: |  |:| || :|   |     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    354 CP 355
          |:
Qy    109 CS 110


RESULT    5
ID   CYR6_MOUSE     STANDARD;      PRT;   379 AA.
AC   P18406;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (3CH61).
GN   IGFBP10 OR CYR61.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=FIBROBLAST;
RX   MEDLINE; 90287146.
RA   O'BRIEN T.P., YANG G.P., SANDERS L., LAU L.F.;
RT   "Expression of cyr61, a growth factor-inducible immediate-early
RT   gene.";
RL   MOL. CELL. BIOL. 10:3569-3577(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AJ; TISSUE=EMBRYONIC FIBROBLAST;
RX   MEDLINE; 91288203.
RA   LATINKIC B.V., O'BRIEN T.P., LAU L.F.;
RT   "Promoter function and structure of the growth factor-inducible
RT   immediate early gene cyr61.";
RL   NUCLEIC ACIDS RES. 19:3261-3267(1991).
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION.
CC   -!- TISSUE SPECIFICITY: LOW IN KIDNEY, ADRENAL GLAND, TESTES, BRAIN,
CC       AND OVARY, MODERATE IN HEART, UTERUS, AND SKELETAL MUSCLE, HIGHEST
CC       IN LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED FROM G(0)/G(1) THROUGH MID-G(1) IN
CC       NORMAL CELLS, AND AT A CONSTANT LEVEL IN RAPIDLY GROWING CELLS.
CC   -!- INDUCTION: BY GROWTH FACTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M32490; G309206; -.
DR   EMBL; X56790; G50633; -.
DR   PIR; A35669; A35669.
DR   MGD; MGI:88613; IGFBP10.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25   379       CYR61 PROTEIN.
FT   DOMAIN      98   164       VWFC.
FT   DOMAIN     284   358       CTCK.
FT   DISULFID   284   321       BY SIMILARITY.
FT   DISULFID   301   335       BY SIMILARITY.
FT   DISULFID   312   351       BY SIMILARITY.
FT   DISULFID   315   353       BY SIMILARITY.
FT   DISULFID   320   357       BY SIMILARITY.
SQ   SEQUENCE   379 AA;  41709 MW;  116B80C7 CRC32;

  Query Match          14.4%;  Score 126;  DB 1;  Length 379;
  Best Local Similarity  32.3%;  Pred. No. 3.54e-09;
  Matches   20;  Conservative   12;  Mismatches  27;  Indels   3;  Gaps   3;

Db    298 YAGCSSVKKYRPKYC-GSCVDGRCCTPLQTRTVKMRFRCEDGEMFSKNVMMIQSCKCNYN 356
          ||:|:: |   : | |:|    || |:::: |  |:| ||  |  :|     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    357 CP 358
          |:
Qy    109 CS 110


RESULT    6
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN   FT.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92069752.
```

```
RA   MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA   GOODMAN C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   CELL 67:853-868(1991).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC       TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC       DIFFERENTIATION AND MORPHOGENESIS, AND DEAT DURING THE PUPAL
CC       STAGE.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY; CONTAINS 37 CADHERINS-TYPE REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M80537; G157409; -.
DR   PIR; A41087; IJFFTM.
DR   FLYBASE; FBgn0001075; ft.
DR   PROSITE; PS00232; CADHERIN; 22.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PFAM; PF00008; EGF; 4.
DR   PFAM; PF00028; cadherin; 33.
DR   PFAM; PF00054; laminin_G; 1.
DR   HSSP; P00740; 1IXA.
KW   CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW   CALCIUM-BINDING; REPEAT; EGF-LIKE DOMAIN.
FT   SIGNAL       1    35       POTENTIAL.
FT   CHAIN       36  5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN      36  4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  4584  4609       POTENTIAL.
FT   DOMAIN    4610  5147       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      36   156       CADHERIN 1.
FT   REPEAT     157   270       CADHERIN 2.
FT   REPEAT     271   382       CADHERIN 3.
FT   REPEAT     383   494       CADHERIN 4.
FT   REPEAT     495   599       CADHERIN 5.
FT   REPEAT     600   708       CADHERIN 6.
FT   REPEAT     709   820       CADHERIN 7.
FT   REPEAT     821   942       CADHERIN 8.
FT   REPEAT     943  1049       CADHERIN 9.
FT   REPEAT    1050  1153       CADHERIN 10.
FT   REPEAT    1154  1278       CADHERIN 11.
FT   REPEAT    1279  1384       CADHERIN 12.
FT   REPEAT    1385  1489       CADHERIN 13.
FT   REPEAT    1490  1601       CADHERIN 14.
FT   REPEAT    1602  1713       CADHERIN 15.
FT   REPEAT    1714  1823       CADHERIN 16.
FT   REPEAT    1824  1922       CADHERIN 17.
FT   REPEAT    1923  2027       CADHERIN 18.
FT   REPEAT    2028  2167       CADHERIN 19.
FT   REPEAT    2168  2278       CADHERIN 20.
FT   REPEAT    2279  2385       CADHERIN 21.
FT   REPEAT    2386  2491       CADHERIN 22.
FT   REPEAT    2492  2596       CADHERIN 23.
FT   REPEAT    2597  2703       CADHERIN 24.
FT   REPEAT    2704  2810       CADHERIN 25.
FT   REPEAT    2811  2913       CADHERIN 26.
FT   REPEAT    2914  3013       CADHERIN 27.
FT   REPEAT    3014  3124       CADHERIN 28.
FT   REPEAT    3125  3229       CADHERIN 29.
FT   REPEAT    3230  3334       CADHERIN 30.
FT   REPEAT    3335  3439       CADHERIN 31.
FT   REPEAT    3440  3545       CADHERIN 32.
FT   REPEAT    3546  3651       CADHERIN 33.
FT   REPEAT    3652  3756       CADHERIN 34.
FT   DOMAIN    3950  4011       EGF-LIKE 1.
FT   DOMAIN    4013  4049       EGF-LIKE 2.
FT   DOMAIN    4052  4090       EGF-LIKE 3.
FT   DOMAIN    4092  4128       EGF-LIKE 4.
FT   DOMAIN    4321  4362       EGF-LIKE 5.
FT   DISULFID  3954  3966       BY SIMILARITY.
FT   DISULFID  3960  3999       BY SIMILARITY.
FT   DISULFID  4001  4010       BY SIMILARITY.
FT   DISULFID  4017  4028       BY SIMILARITY.
FT   DISULFID  4022  4037       BY SIMILARITY.
FT   DISULFID  4039  4048       BY SIMILARITY.
FT   DISULFID  4056  4067       BY SIMILARITY.
FT   DISULFID  4061  4078       BY SIMILARITY.
FT   DISULFID  4080  4089       BY SIMILARITY.
FT   DISULFID  4096  4107       BY SIMILARITY.
FT   DISULFID  4101  4116       BY SIMILARITY.
FT   DISULFID  4118  4127       BY SIMILARITY.
FT   DISULFID  4325  4341       BY SIMILARITY.
FT   DISULFID  4334  4350       BY SIMILARITY.
FT   DISULFID  4352  4361       BY SIMILARITY.
FT   CARBOHYD   239   239       POTENTIAL.
FT   CARBOHYD   257   257       POTENTIAL.
FT   CARBOHYD   276   276       POTENTIAL.
FT   CARBOHYD   280   280       POTENTIAL.
FT   CARBOHYD   402   402       POTENTIAL.
FT   CARBOHYD   461   461       POTENTIAL.
FT   CARBOHYD   605   605       POTENTIAL.
FT   CARBOHYD   631   631       POTENTIAL.
FT   CARBOHYD  1155  1155       POTENTIAL.
FT   CARBOHYD  1367  1367       POTENTIAL.
FT   CARBOHYD  1458  1458       POTENTIAL.
FT   CARBOHYD  1751  1751       POTENTIAL.
FT   CARBOHYD  1831  1831       POTENTIAL.
FT   CARBOHYD  1880  1880       POTENTIAL.
FT   CARBOHYD  2080  2080       POTENTIAL.
FT   CARBOHYD  2171  2171       POTENTIAL.
FT   CARBOHYD  2247  2247       POTENTIAL.
FT   CARBOHYD  2290  2290       POTENTIAL.
FT   CARBOHYD  2437  2437       POTENTIAL.
FT   CARBOHYD  2581  2581       POTENTIAL.
FT   CARBOHYD  2799  2799       POTENTIAL.
FT   CARBOHYD  2920  2920       POTENTIAL.
FT   CARBOHYD  2946  2946       POTENTIAL.
FT   CARBOHYD  2967  2967       POTENTIAL.
FT   CARBOHYD  3167  3167       POTENTIAL.
FT   CARBOHYD  3303  3303       POTENTIAL.
FT   CARBOHYD  3386  3386       POTENTIAL.
FT   CARBOHYD  3389  3389       POTENTIAL.
FT   CARBOHYD  3525  3525       POTENTIAL.
FT   CARBOHYD  3852  3852       POTENTIAL.
FT   CARBOHYD  3865  3865       POTENTIAL.
FT   CARBOHYD  3905  3905       POTENTIAL.
FT   CARBOHYD  4306  4306       POTENTIAL.
FT   CARBOHYD  4414  4414       POTENTIAL.
FT   CARBOHYD  4471  4471       POTENTIAL.
FT   CARBOHYD  4487  4487       POTENTIAL.
FT   CARBOHYD  4539  4539       POTENTIAL.
FT   CARBOHYD  4550  4550       POTENTIAL.
FT   VARIANT   1229  1229       S -> G.
FT   VARIANT   1233  1233       G -> S.
SQ   SEQUENCE  5147 AA;  564868 MW;  1EF20E13 CRC32;

  Query Match          14.4%;  Score 126;  DB 1;  Length 5147;
  Best Local Similarity  44.4%;  Pred. No. 3.54e-09;
  Matches   16;  Conservative   9;  Mismatches   8;  Indels   3;  Gaps   3;

Db   4061 CRNGGSCQRSPDGSSYFCLCRPGFRGNQCESVSDSC 4096
          |::|  |: |  | :| |||:||| |::||   ::|
Qy      4 CHHG-QCHISDRGEPY-CLCQPGFSGHHCEQ-ENPC 36
```

```
RESULT   7
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1     90
FT   CHAIN       91   2139       CRUMBS PROTEIN.
FT   DOMAIN      91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085   2111       POTENTIAL.
FT   DOMAIN    2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267    303       EGF-LIKE 1.
FT   DOMAIN     306    343       EGF-LIKE 2.
FT   DOMAIN     348    386       EGF-LIKE 3.
FT   DOMAIN     388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427    463       EGF-LIKE 5.
FT   DOMAIN     464    500       EGF-LIKE 6.
FT   DOMAIN     501    532       EGF-LIKE 7.
FT   DOMAIN     545    581       EGF-LIKE 8.
FT   DOMAIN     582    611       EGF-LIKE 9.
FT   DOMAIN     609    646       EGF-LIKE 10.
FT   DOMAIN     648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840    902       EGF-LIKE 16.
FT   DOMAIN     904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980   1021       EGF-LIKE 19.
FT   DOMAIN    1207   1243       EGF-LIKE 20.
FT   DOMAIN    1481   1517       EGF-LIKE 21.
FT   DOMAIN    1759   1795       EGF-LIKE 22.
FT   DOMAIN    1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874   1915       EGF-LIKE 25.
FT   DOMAIN    1915   1951       EGF-LIKE 26.
FT   DOMAIN    1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030   2070       EGF-LIKE 29.
FT   DISULFID   271    282       BY SIMILARITY.
FT   DISULFID   276    291       BY SIMILARITY.
FT   DISULFID   293    302       BY SIMILARITY.
FT   DISULFID   310    321       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   333    342       BY SIMILARITY.
FT   DISULFID   352    363       BY SIMILARITY.
FT   DISULFID   357    374       BY SIMILARITY.
FT   DISULFID   376    385       BY SIMILARITY.
FT   DISULFID   392    403       BY SIMILARITY.
FT   DISULFID   397    412       BY SIMILARITY.
FT   DISULFID   414    424       BY SIMILARITY.
FT   DISULFID   431    442       BY SIMILARITY.
FT   DISULFID   436    451       BY SIMILARITY.
FT   DISULFID   453    462       BY SIMILARITY.
FT   DISULFID   468    479       BY SIMILARITY.
FT   DISULFID   473    488       BY SIMILARITY.
FT   DISULFID   490    499       BY SIMILARITY.
FT   DISULFID   505    515       BY SIMILARITY.
FT   DISULFID   509    520       BY SIMILARITY.
FT   DISULFID   522    531       BY SIMILARITY.
FT   DISULFID   549    562       BY SIMILARITY.
FT   DISULFID   556    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   586    597       BY SIMILARITY.
FT   DISULFID   591    602       BY SIMILARITY.
FT   DISULFID   604    610       BY SIMILARITY.
FT   DISULFID   613    624       BY SIMILARITY.
FT   DISULFID   618    634       BY SIMILARITY.
FT   DISULFID   636    645       BY SIMILARITY.
FT   DISULFID   652    664       BY SIMILARITY.
FT   DISULFID   659    673       BY SIMILARITY.
FT   DISULFID   675    684       BY SIMILARITY.
FT   DISULFID   691    702       BY SIMILARITY.
FT   DISULFID   696    711       BY SIMILARITY.
FT   DISULFID   713    722       BY SIMILARITY.
FT   DISULFID   729    740       BY SIMILARITY.
FT   DISULFID   734    749       BY SIMILARITY.
FT   DISULFID   751    760       BY SIMILARITY.
FT   DISULFID   767    778       BY SIMILARITY.
FT   DISULFID   772    787       BY SIMILARITY.
FT   DISULFID   789    799       BY SIMILARITY.
FT   DISULFID   806    817       BY SIMILARITY.
FT   DISULFID   811    826       BY SIMILARITY.
FT   DISULFID   828    837       BY SIMILARITY.
FT   DISULFID   844    855       BY SIMILARITY.
FT   DISULFID   849    890       BY SIMILARITY.
FT   DISULFID   892    901       BY SIMILARITY.
FT   DISULFID   908    919       BY SIMILARITY.
FT   DISULFID   913    928       BY SIMILARITY.
FT   DISULFID   930    939       BY SIMILARITY.
FT   DISULFID   946    957       BY SIMILARITY.
FT   DISULFID   952    966       BY SIMILARITY.
```

```
FT   DISULFID   968   977       BY SIMILARITY.
FT   DISULFID   984   995       BY SIMILARITY.
FT   DISULFID   989  1009       BY SIMILARITY.
FT   DISULFID  1011  1020       BY SIMILARITY.
FT   DISULFID  1211  1222       BY SIMILARITY.
FT   DISULFID  1216  1231       BY SIMILARITY.
FT   DISULFID  1233  1242       BY SIMILARITY.
FT   DISULFID  1485  1496       BY SIMILARITY.
FT   DISULFID  1490  1505       BY SIMILARITY.
FT   DISULFID  1507  1516       BY SIMILARITY.
FT   DISULFID  1763  1774       BY SIMILARITY.
FT   DISULFID  1768  1783       BY SIMILARITY.
FT   DISULFID  1785  1794       BY SIMILARITY.
FT   DISULFID  1801  1812       BY SIMILARITY.
FT   DISULFID  1806  1821       BY SIMILARITY.
FT   DISULFID  1823  1832       BY SIMILARITY.
FT   DISULFID  1839  1850       BY SIMILARITY.
FT   DISULFID  1844  1859       BY SIMILARITY.
FT   DISULFID  1861  1870       BY SIMILARITY.
FT   DISULFID  1878  1889       BY SIMILARITY.
FT   DISULFID  1883  1903       BY SIMILARITY.
FT   DISULFID  1905  1914       BY SIMILARITY.
FT   DISULFID  1919  1930       BY SIMILARITY.
FT   DISULFID  1924  1939       BY SIMILARITY.
FT   DISULFID  1941  1950       BY SIMILARITY.
FT   DISULFID  1957  1968       BY SIMILARITY.
FT   DISULFID  1962  1977       BY SIMILARITY.
FT   DISULFID  1979  1988       BY SIMILARITY.
FT   DISULFID  1995  2008       BY SIMILARITY.
FT   DISULFID  2002  2017       BY SIMILARITY.
FT   DISULFID  2019  2028       BY SIMILARITY.
FT   CARBOHYD    37    37       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   198   198       POTENTIAL.
FT   CARBOHYD   238   238       POTENTIAL.
FT   CARBOHYD   239   239       POTENTIAL.
FT   CARBOHYD   336   336       POTENTIAL.
FT   CARBOHYD   400   400       POTENTIAL.
FT   CARBOHYD   550   550       POTENTIAL.
FT   CARBOHYD   565   565       POTENTIAL.
FT   CARBOHYD   736   736       POTENTIAL.
FT   CARBOHYD   746   746       POTENTIAL.
FT   CARBOHYD   860   860       POTENTIAL.
FT   CARBOHYD   884   884       POTENTIAL.
FT   CARBOHYD   976   976       POTENTIAL.
FT   CARBOHYD  1102  1102       POTENTIAL.
FT   CARBOHYD  1114  1114       POTENTIAL.
FT   CARBOHYD  1138  1138       POTENTIAL.
FT   CARBOHYD  1192  1192       POTENTIAL.
FT   CARBOHYD  1245  1245       POTENTIAL.
FT   CARBOHYD  1255  1255       POTENTIAL.
FT   CARBOHYD  1354  1354       POTENTIAL.
FT   CARBOHYD  1363  1363       POTENTIAL.
FT   CARBOHYD  1441  1441       POTENTIAL.
FT   CARBOHYD  1454  1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          14.3%;  Score 125;  DB 1;  Length 2139;
  Best Local Similarity  42.1%;  Pred. No. 5.62e-09;
  Matches   16;  Conservative   10;  Mismatches   9;  Indels   3;  Gaps   3;

Db   1806 CLNNGTC-INQVAAFFCQCQPGFEGQHCEQNIDECADQ 1842
          | ::| | |:: :  :| ||||| |:||||: : |  :
Qy      4 C-HHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT    8
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19  >2444       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19  1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737  1757       POTENTIAL.
FT   DOMAIN    1758  >2444       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20    58       EGF-LIKE 1.
FT   DOMAIN      59    99       EGF-LIKE 2.
FT   DOMAIN     102   139       EGF-LIKE 3.
FT   DOMAIN     140   176       EGF-LIKE 4.
FT   DOMAIN     178   216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218   255       EGF-LIKE 6.
FT   DOMAIN     257   293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295   333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335   371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372   410       EGF-LIKE 10.
FT   DOMAIN     412   450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452   488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490   526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528   564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566   601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603   639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641   676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678   714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716   751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753   789       EGF-LIKE 20.
FT   DOMAIN     791   827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829   868       EGF-LIKE 22.
```

```
FT   DOMAIN      870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      908    944       EGF-LIKE 24.
FT   DOMAIN      946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984   1020       EGF-LIKE 26.
FT   DOMAIN     1022   1058       EGF-LIKE 27.
FT   DOMAIN     1060   1096       EGF-LIKE 28.
FT   DOMAIN     1098   1144       EGF-LIKE 29.
FT   DOMAIN     1146   1182       EGF-LIKE 30.
FT   DOMAIN     1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1268   1306       EGF-LIKE 33.
FT   DOMAIN     1308   1347       EGF-LIKE 34.
FT   DOMAIN     1349   1385       EGF-LIKE 35.
FT   DOMAIN     1388   1427       EGF-LIKE 36.
FT   DOMAIN     1446   1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1481       LIN/NOTCH 1.
FT   REPEAT     1482   1523       LIN/NOTCH 2.
FT   REPEAT     1524   1563       LIN/NOTCH 3.
FT   DOMAIN     1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT     1876   1921       ANK MOTIF 1.
FT   REPEAT     1923   1954       ANK MOTIF 2.
FT   REPEAT     1956   1987       ANK MOTIF 3.
FT   REPEAT     1990   2021       ANK MOTIF 4.
FT   REPEAT     2023   2054       ANK MOTIF 5.
FT   REPEAT     2056   2087       ANK MOTIF 6.
FT   DOMAIN     1576   1579       POLY-VAL.
FT   DOMAIN     1662   1665       POLY-ARG.
FT   DOMAIN     1729   1732       POLY-PRO.
FT   DOMAIN     1741   1744       POLY-ALA.
FT   DOMAIN     1902   1905       POLY-GLU.
FT   DOMAIN     2260   2263       POLY-GLY.
FT   DOMAIN     2404   2407       POLY-GLN.
FT   DOMAIN     2411   2418       POLY-PRO.
FT   DISULFID     24     37       BY SIMILARITY.
FT   DISULFID     31     46       BY SIMILARITY.
FT   DISULFID     48     57       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    261    272       BY SIMILARITY.
FT   DISULFID    266    281       BY SIMILARITY.
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    299    312       BY SIMILARITY.
FT   DISULFID    306    321       BY SIMILARITY.
FT   DISULFID    323    332       BY SIMILARITY.
FT   DISULFID    339    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    361    370       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    398       BY SIMILARITY.
FT   DISULFID    400    409       BY SIMILARITY.
FT   DISULFID    416    429       BY SIMILARITY.
FT   DISULFID    423    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    456    467       BY SIMILARITY.
FT   DISULFID    461    476       BY SIMILARITY.
FT   DISULFID    478    487       BY SIMILARITY.
FT   DISULFID    494    505       BY SIMILARITY.
FT   DISULFID    499    514       BY SIMILARITY.
FT   DISULFID    516    525       BY SIMILARITY.
FT   DISULFID    532    543       BY SIMILARITY.
FT   DISULFID    537    552       BY SIMILARITY.
FT   DISULFID    554    563       BY SIMILARITY.
FT   DISULFID    570    580       BY SIMILARITY.
FT   DISULFID    575    589       BY SIMILARITY.
FT   DISULFID    591    600       BY SIMILARITY.
FT   DISULFID    607    618       BY SIMILARITY.
FT   DISULFID    612    627       BY SIMILARITY.
FT   DISULFID    629    638       BY SIMILARITY.
FT   DISULFID    645    655       BY SIMILARITY.
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    867       BY SIMILARITY.
FT   DISULFID    874    885       BY SIMILARITY.
FT   DISULFID    879    894       BY SIMILARITY.
FT   DISULFID    896    905       BY SIMILARITY.
FT   DISULFID    912    923       BY SIMILARITY.
FT   DISULFID    917    932       BY SIMILARITY.
FT   DISULFID    934    943       BY SIMILARITY.
FT   DISULFID    988    999       BY SIMILARITY.
FT   DISULFID    993   1008       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1026   1037       BY SIMILARITY.
FT   DISULFID   1031   1046       BY SIMILARITY.
FT   DISULFID   1048   1057       BY SIMILARITY.
FT   DISULFID   1064   1075       BY SIMILARITY.
FT   DISULFID   1069   1084       BY SIMILARITY.
FT   DISULFID   1086   1095       BY SIMILARITY.
FT   DISULFID   1102   1123       BY SIMILARITY.
FT   DISULFID   1117   1132       BY SIMILARITY.
FT   DISULFID   1134   1143       BY SIMILARITY.
FT   DISULFID   1150   1161       BY SIMILARITY.
FT   DISULFID   1155   1170       BY SIMILARITY.
FT   DISULFID   1172   1181       BY SIMILARITY.
FT   DISULFID   1188   1199       BY SIMILARITY.
FT   DISULFID   1193   1208       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          14.3%;  Score 125;  DB 1;  Length 2444;
  Best Local Similarity  23.7%;  Pred. No. 5.62e-09;
  Matches   27;  Conservative   32;  Mismatches  46;  Indels   9;  Gaps   8;

Db    575 CHYGSCK-DGVATFTCLCRPGYTGHHCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGT 633
          || | |     :   |||:||::||||| : | | ::   :| : :  :: |   |  ::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE--IVREAIR-RQKDYAS-CATAS 58

Db    634 TGPNCEINLDDCASSPCDSGTCLDKIDGYECACEPGYTGSMCNSNIDECAGNPC 687
            | |     |::: |::     :   ::|: :::   : : :: ||:  :|
Qy     59 KVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT    9
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
```

```
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
FT   DOMAIN     869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907    943       EGF-LIKE 24.
FT   DOMAIN     945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983   1019       EGF-LIKE 26.
FT   DOMAIN    1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059   1095       EGF-LIKE 28.
FT   DOMAIN    1097   1143       EGF-LIKE 29.
FT   DOMAIN    1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267   1305       EGF-LIKE 33.
FT   DOMAIN    1307   1346       EGF-LIKE 34.
FT   DOMAIN    1348   1384       EGF-LIKE 35.
FT   DOMAIN    1387   1426       EGF-LIKE 36.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1865   2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2076       ANK MOTIF 6.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
FT   DISULFID   339    350       BY SIMILARITY.
FT   DISULFID   344    359       BY SIMILARITY.
FT   DISULFID   361    370       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    398       BY SIMILARITY.
FT   DISULFID   400    409       BY SIMILARITY.
FT   DISULFID   416    429       BY SIMILARITY.
FT   DISULFID   423    438       BY SIMILARITY.
FT   DISULFID   440    449       BY SIMILARITY.
FT   DISULFID   456    467       BY SIMILARITY.
FT   DISULFID   461    476       BY SIMILARITY.
FT   DISULFID   478    487       BY SIMILARITY.
FT   DISULFID   494    505       BY SIMILARITY.
FT   DISULFID   499    514       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   532    543       BY SIMILARITY.
FT   DISULFID   537    552       BY SIMILARITY.
FT   DISULFID   554    563       BY SIMILARITY.
FT   DISULFID   570    580       BY SIMILARITY.
FT   DISULFID   575    589       BY SIMILARITY.
FT   DISULFID   591    600       BY SIMILARITY.
FT   DISULFID   607    618       BY SIMILARITY.
FT   DISULFID   612    627       BY SIMILARITY.
FT   DISULFID   629    638       BY SIMILARITY.
FT   DISULFID   645    655       BY SIMILARITY.
FT   DISULFID   650    664       BY SIMILARITY.
FT   DISULFID   666    675       BY SIMILARITY.
FT   DISULFID   682    693       BY SIMILARITY.
FT   DISULFID   687    702       BY SIMILARITY.
FT   DISULFID   704    713       BY SIMILARITY.
FT   DISULFID   720    730       BY SIMILARITY.
FT   DISULFID   725    739       BY SIMILARITY.
```

```
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    866       BY SIMILARITY.
FT   DISULFID    873    884       BY SIMILARITY.
FT   DISULFID    878    893       BY SIMILARITY.
FT   DISULFID    895    904       BY SIMILARITY.
FT   DISULFID    911    922       BY SIMILARITY.
FT   DISULFID    916    931       BY SIMILARITY.
FT   DISULFID    933    942       BY SIMILARITY.
FT   DISULFID    987    998       BY SIMILARITY.
FT   DISULFID    992   1007       BY SIMILARITY.
FT   DISULFID   1009   1018       BY SIMILARITY.
FT   DISULFID   1025   1036       BY SIMILARITY.
FT   DISULFID   1030   1045       BY SIMILARITY.
FT   DISULFID   1047   1056       BY SIMILARITY.
FT   DISULFID   1063   1074       BY SIMILARITY.
FT   DISULFID   1068   1083       BY SIMILARITY.
FT   DISULFID   1085   1094       BY SIMILARITY.
FT   DISULFID   1101   1122       BY SIMILARITY.
FT   DISULFID   1116   1131       BY SIMILARITY.
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.
FT   DISULFID   1225   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1271   1284       BY SIMILARITY.
FT   DISULFID   1276   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1311   1322       BY SIMILARITY.
FT   DISULFID   1316   1334       BY SIMILARITY.
FT   DISULFID   1336   1345       BY SIMILARITY.
FT   DISULFID   1352   1363       BY SIMILARITY.
FT   DISULFID   1357   1372       BY SIMILARITY.
FT   DISULFID   1374   1383       BY SIMILARITY.
FT   DISULFID   1391   1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          14.3%;  Score 125;  DB 1;  Length 2531;
  Best Local Similarity  25.4%;  Pred. No. 5.62e-09;
  Matches   30;  Conservative   31;  Mismatches  45;  Indels  12;  Gaps  11;

Db    570 CDPDPCHIGLCKDGVATFTCLCQPGYTGHHCETNINECHSQPCRHGGTCQDRDNYYLCLC 629
          |   |||:  : |   : ||||||::||||| : | | ::  | :   :::|  |   |
Qy      4 CHHGQCHISD-R-G-EPY-CLCQPGFSGHHCEQE-NPCMGEIVREA-IRRQKD--YAS-C 54

Db    630 LKGTTGPNCEINLDDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNVNIDECAGSPC 687
            ::  | |     |::  |::     :   ::|:  :::   : : :: ||:  :|
Qy     55 ATASKVPIMECR-GGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEV-ERHL-ECGCRAC 109


RESULT  10
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
```

```
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     44       POTENTIAL.
FT   CHAIN       45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746   1766       POTENTIAL.
FT   DOMAIN    1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58     95       EGF-LIKE 1.
FT   DOMAIN      96    136       EGF-LIKE 2.
FT   DOMAIN     139    176       EGF-LIKE 3.
FT   DOMAIN     177    215       EGF-LIKE 4.
FT   DOMAIN     217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255    291       EGF-LIKE 6.
FT   DOMAIN     293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
     DOMAIN     372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
     DOMAIN     409    447       EGF-LIKE 10.
     DOMAIN     449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     867    905       EGF-LIKE 22.
FT   DOMAIN     907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 25.
FT   DOMAIN    1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1060   1096       EGF-LIKE 27.
FT   DOMAIN    1098   1134       EGF-LIKE 28.
FT   DOMAIN    1136   1181       EGF-LIKE 29.
FT   DOMAIN    1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1297   1335       EGF-LIKE 33.
FT   DOMAIN    1337   1373       EGF-LIKE 34.
FT   DOMAIN    1375   1412       EGF-LIKE 35.
FT   DOMAIN    1415   1451       EGF-LIKE 36.
     DOMAIN    1475   1593       3 X LIN/NOTCH REPEATS.
     REPEAT    1475   1513       LIN/NOTCH 1.
     REPEAT    1514   1553       LIN/NOTCH 2.
FT   REPEAT    1554   1593       LIN/NOTCH 3.
FT   DOMAIN    1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN    2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     83       BY SIMILARITY.
FT   DISULFID    85     94       BY SIMILARITY.
FT   DISULFID   100    111       BY SIMILARITY.
FT   DISULFID   105    124       BY SIMILARITY.
FT   DISULFID   126    135       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   181    192       BY SIMILARITY.
FT   DISULFID   186    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    241       BY SIMILARITY.
FT   DISULFID   243    252       BY SIMILARITY.
FT   DISULFID   259    270       BY SIMILARITY.
FT   DISULFID   264    279       BY SIMILARITY.
FT   DISULFID   281    290       BY SIMILARITY.
FT   DISULFID   297    308       BY SIMILARITY.
FT   DISULFID   302    317       BY SIMILARITY.
FT   DISULFID   319    328       BY SIMILARITY.
FT   DISULFID   335    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    396       BY SIMILARITY.
FT   DISULFID   398    407       BY SIMILARITY.
FT   DISULFID   413    424       BY SIMILARITY.
FT   DISULFID   418    435       BY SIMILARITY.
FT   DISULFID   437    446       BY SIMILARITY.
FT   DISULFID   453    465       BY SIMILARITY.
FT   DISULFID   459    474       BY SIMILARITY.
FT   DISULFID   476    485       BY SIMILARITY.
FT   DISULFID   492    503       BY SIMILARITY.
FT   DISULFID   497    512       BY SIMILARITY.
FT   DISULFID   514    523       BY SIMILARITY.
FT   DISULFID   530    541       BY SIMILARITY.
FT   DISULFID   535    550       BY SIMILARITY.
FT   DISULFID   552    561       BY SIMILARITY.
FT   DISULFID   568    579       BY SIMILARITY.
FT   DISULFID   573    588       BY SIMILARITY.
FT   DISULFID   590    599       BY SIMILARITY.
FT   DISULFID   606    616       BY SIMILARITY.
FT   DISULFID   611    625       BY SIMILARITY.
FT   DISULFID   627    636       BY SIMILARITY.
FT   DISULFID   643    654       BY SIMILARITY.
FT   DISULFID   648    663       BY SIMILARITY.
FT   DISULFID   665    674       BY SIMILARITY.
FT   DISULFID   681    692       BY SIMILARITY.
FT   DISULFID   686    701       BY SIMILARITY.
FT   DISULFID   703    712       BY SIMILARITY.
FT   DISULFID   719    730       BY SIMILARITY.
FT   DISULFID   724    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    853       BY SIMILARITY.
FT   DISULFID   855    864       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          14.0%;  Score 122;  DB 1;  Length 2703;
  Best Local Similarity  40.5%;  Pred. No. 2.23e-08;
  Matches   17;  Conservative    9;  Mismatches  11;  Indels   5;  Gaps   5;

Db    952 SFPCQNGGTC-LDGIGD-YSCLCVDGFDGKHCETDINECLSQ 991
          :| |::|  | :   |: .| |||  || | ||| : | |:::
Qy      1 AFKCHHG-QCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGE 39


RESULT  11
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
```

```
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. VIROL. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG, AND TESTIS;
RX   MEDLINE; 96281668.
RA   UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   DEVELOPMENT 122:2251-2259(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; G1714084; -.
DR   EMBL; U43691; G1401160; -.
DR   PIR; A38072; TVMVT3.
DR   MGD; MGI:107471; NOTCH4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   DOMAIN      21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1444   1464       POTENTIAL.
FT   DOMAIN    1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     60       EGF-LIKE 1.
FT   DOMAIN      61    112       EGF-LIKE 2.
FT   DOMAIN     115    152       EGF-LIKE 3.
FT   DOMAIN     153    189       EGF-LIKE 4.
FT   DOMAIN     191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     231    271       EGF-LIKE 6.
FT   DOMAIN     273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     389    427       EGF-LIKE 10.
FT   DOMAIN     429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     622    656       EGF-LIKE 16.
FT   DOMAIN     658    686       EGF-LIKE 17.
FT   DOMAIN     688    724       EGF-LIKE 18.
FT   DOMAIN     726    762       EGF-LIKE 19.
FT   DOMAIN     764    800       EGF-LIKE 20.
FT   DOMAIN     803    839       EGF-LIKE 21.
FT   DOMAIN     841    877       EGF-LIKE 22.
FT   DOMAIN     878    924       EGF-LIKE 23.
FT   DOMAIN     926    962       EGF-LIKE 24.
FT   DOMAIN     964   1000       EGF-LIKE 25.
FT   DOMAIN    1002   1040       EGF-LIKE 26.
FT   DOMAIN    1042   1081       EGF-LIKE 27.
FT   DOMAIN    1083   1122       EGF-LIKE 28.
FT   DOMAIN    1126   1167       EGF-LIKE 29.
FT   DOMAIN    1168   1282       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1168   1208       LIN/NOTCH 1.
FT   REPEAT    1209   1242       LIN/NOTCH 2.
FT   REPEAT    1243   1282       LIN/NOTCH 3.
FT   DOMAIN    1572   1785       6 X ANK MOTIF REPEATS.
FT   REPEAT    1572   1603       ANK MOTIF 1.
FT   REPEAT    1622   1653       ANK MOTIF 2.
FT   REPEAT    1654   1685       ANK MOTIF 3.
FT   REPEAT    1688   1719       ANK MOTIF 4.
FT   REPEAT    1721   1752       ANK MOTIF 5.
FT   REPEAT    1754   1785       ANK MOTIF 6.
FT   DISULFID    25     38       BY SIMILARITY.
FT   DISULFID    32     48       BY SIMILARITY.
FT   DISULFID    50     59       BY SIMILARITY.
FT   DISULFID    65     77       BY SIMILARITY.
FT   DISULFID    71    100       BY SIMILARITY.
FT   DISULFID   102    111       BY SIMILARITY.
FT   DISULFID   119    130       BY SIMILARITY.
FT   DISULFID   124    140       BY SIMILARITY.
FT   DISULFID   142    151       BY SIMILARITY.
FT   DISULFID   157    168       BY SIMILARITY.
FT   DISULFID   162    177       BY SIMILARITY.
FT   DISULFID   179    188       BY SIMILARITY.
FT   DISULFID   195    208       BY SIMILARITY.
FT   DISULFID   202    217       BY SIMILARITY.
FT   DISULFID   219    228       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   235    246       BY SIMILARITY.
FT   DISULFID   240    259       BY SIMILARITY.
FT   DISULFID   261    270       BY SIMILARITY.
FT   DISULFID   277    288       BY SIMILARITY.
FT   DISULFID   282    297       BY SIMILARITY.
FT   DISULFID   299    308       BY SIMILARITY.
FT   DISULFID   315    329       BY SIMILARITY.
FT   DISULFID   323    338       BY SIMILARITY.
FT   DISULFID   340    349       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    376       BY SIMILARITY.
FT   DISULFID   378    387       BY SIMILARITY.
FT   DISULFID   393    404       BY SIMILARITY.
FT   DISULFID   398    415       BY SIMILARITY.
FT   DISULFID   417    426       BY SIMILARITY.
FT   DISULFID   433    449       BY SIMILARITY.
FT   DISULFID   443    458       BY SIMILARITY.
FT   DISULFID   460    469       BY SIMILARITY.
FT   DISULFID   476    487       BY SIMILARITY.
FT   DISULFID   481    496       BY SIMILARITY.
FT   DISULFID   498    507       BY SIMILARITY.
FT   DISULFID   514    525       BY SIMILARITY.
FT   DISULFID   519    534       BY SIMILARITY.
FT   DISULFID   536    545       BY SIMILARITY.
FT   DISULFID   552    563       BY SIMILARITY.
FT   DISULFID   557    572       BY SIMILARITY.
FT   DISULFID   574    583       BY SIMILARITY.
FT   DISULFID   590    601       BY SIMILARITY.
FT   DISULFID   595    610       BY SIMILARITY.
FT   DISULFID   612    621       BY SIMILARITY.
FT   DISULFID   626    637       BY SIMILARITY.
FT   DISULFID   631    646       BY SIMILARITY.
FT   DISULFID   648    655       BY SIMILARITY.
FT   DISULFID   662    669       BY SIMILARITY.
FT   DISULFID   664    674       BY SIMILARITY.
FT   DISULFID   676    685       BY SIMILARITY.
FT   DISULFID   692    703       BY SIMILARITY.
FT   DISULFID   697    712       BY SIMILARITY.
FT   DISULFID   714    723       BY SIMILARITY.
```

```
FT   DISULFID   730   741       BY SIMILARITY.
FT   DISULFID   735   750       BY SIMILARITY.
FT   DISULFID   752   761       BY SIMILARITY.
FT   DISULFID   768   779       BY SIMILARITY.
FT   DISULFID   773   788       BY SIMILARITY.
FT   DISULFID   790   799       BY SIMILARITY.
FT   DISULFID   807   818       BY SIMILARITY.
FT   DISULFID   812   827       BY SIMILARITY.
FT   DISULFID   829   838       BY SIMILARITY.
FT   DISULFID   845   856       BY SIMILARITY.
FT   DISULFID   850   865       BY SIMILARITY.
FT   DISULFID   867   876       BY SIMILARITY.
FT   DISULFID   882   903       BY SIMILARITY.
FT   DISULFID   897   912       BY SIMILARITY.
FT   DISULFID   914   923       BY SIMILARITY.
FT   DISULFID   930   941       BY SIMILARITY.
FT   DISULFID   935   950       BY SIMILARITY.
FT   DISULFID   952   961       BY SIMILARITY.
FT   DISULFID   968   979       BY SIMILARITY.
FT   DISULFID   973   988       BY SIMILARITY.
FT   DISULFID   990   999       BY SIMILARITY.
FT   DISULFID  1006  1019       BY SIMILARITY.
FT   DISULFID  1011  1028       BY SIMILARITY.
FT   DISULFID  1030  1039       BY SIMILARITY.
FT   DISULFID  1046  1057       BY SIMILARITY.
FT   DISULFID  1051  1069       BY SIMILARITY.
FT   DISULFID  1071  1080       BY SIMILARITY.
FT   DISULFID  1087  1098       BY SIMILARITY.
FT   DISULFID  1092  1110       BY SIMILARITY.
FT   DISULFID  1112  1121       BY SIMILARITY.
FT   DISULFID  1130  1142       BY SIMILARITY.
FT   DISULFID  1136  1155       BY SIMILARITY.
FT   DISULFID  1157  1166       BY SIMILARITY.
FT   CARBOHYD   711   711       POTENTIAL.
FT   CARBOHYD   960   960       POTENTIAL.
FT   CARBOHYD  1139  1139       POTENTIAL.
FT   CONFLICT    43    43       Q -> R (IN REF. 3).
FT   CONFLICT   298   298       L -> P (IN REF. 3).
FT   CONFLICT   884   884       M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match          13.9%;  Score 121;  DB 1;  Length 1964;
  Best Local Similarity  40.5%;  Pred. No. 3.52e-08;
  Matches   15;  Conservative    8;  Mismatches  13;  Indels   1;  Gaps   1;

Db    443 CEHGGSCINTPGSFNCLCLPGYTGSRCEADHNECLSQ 479
          | ||   |: |   ||| ||::| :|| : | |:::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT  12
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20   1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729   1750       POTENTIAL.
FT   DOMAIN    1751   2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     57       EGF-LIKE 1.
FT   DOMAIN      58     99       EGF-LIKE 2.
FT   DOMAIN     102    140       EGF-LIKE 3.
FT   DOMAIN     141    177       EGF-LIKE 4.
FT   DOMAIN     179    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527    563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565    600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     640    675       EGF-LIKE 17.
FT   DOMAIN     677    713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     752    788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     790    826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     828    866       EGF-LIKE 22.
FT   DOMAIN     868    904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     906    942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     944    980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     982   1018       EGF-LIKE 26.
FT   DOMAIN    1020   1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1058   1094       EGF-LIKE 28.
FT   DOMAIN    1096   1142       EGF-LIKE 29.
FT   DOMAIN    1144   1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1182   1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1220   1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1266   1304       EGF-LIKE 33.
FT   DOMAIN    1306   1346       EGF-LIKE 34.
FT   DOMAIN    1347   1383       EGF-LIKE 35.
FT   DOMAIN    1386   1424       EGF-LIKE 36.
FT   DOMAIN    1441   1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1441   1478       LIN/NOTCH 1.
FT   REPEAT    1479   1520       LIN/NOTCH 2.
FT   REPEAT    1521   1560       LIN/NOTCH 3.
FT   DOMAIN    1871   2083       6 X ANK MOTIF REPEATS.
FT   DISULFID    22     35       BY SIMILARITY.
```

```
FT   DISULFID    29     45       BY SIMILARITY.
FT   DISULFID    47     56       BY SIMILARITY.
FT   DISULFID    62     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    128       BY SIMILARITY.
FT   DISULFID   130    139       BY SIMILARITY.
FT   DISULFID   145    156       BY SIMILARITY.
FT   DISULFID   150    165       BY SIMILARITY.
FT   DISULFID   167    176       BY SIMILARITY.
FT   DISULFID   183    194       BY SIMILARITY.
FT   DISULFID   188    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    242       BY SIMILARITY.
FT   DISULFID   244    253       BY SIMILARITY.
FT   DISULFID   260    271       BY SIMILARITY.
FT   DISULFID   265    280       BY SIMILARITY.
FT   DISULFID   282    291       BY SIMILARITY.
FT   DISULFID   298    311       BY SIMILARITY.
FT   DISULFID   305    320       BY SIMILARITY.
FT   DISULFID   322    331       BY SIMILARITY.
FT   DISULFID   338    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   375    386       BY SIMILARITY.
FT   DISULFID   380    397       BY SIMILARITY.
FT   DISULFID   399    408       BY SIMILARITY.
FT   DISULFID   415    428       BY SIMILARITY.
FT   DISULFID   422    437       BY SIMILARITY.
FT   DISULFID   439    448       BY SIMILARITY.
FT   DISULFID   455    466       BY SIMILARITY.
FT   DISULFID   460    475       BY SIMILARITY.
FT   DISULFID   477    486       BY SIMILARITY.
FT   DISULFID   493    504       BY SIMILARITY.
FT   DISULFID   498    513       BY SIMILARITY.
FT   DISULFID   515    524       BY SIMILARITY.
FT   DISULFID   531    542       BY SIMILARITY.
FT   DISULFID   536    551       BY SIMILARITY.
FT   DISULFID   553    562       BY SIMILARITY.
FT   DISULFID   569    579       BY SIMILARITY.
FT   DISULFID   574    588       BY SIMILARITY.
FT   DISULFID   590    599       BY SIMILARITY.
FT   DISULFID   606    617       BY SIMILARITY.
FT   DISULFID   611    626       BY SIMILARITY.
FT   DISULFID   628    637       BY SIMILARITY.
FT   DISULFID   644    654       BY SIMILARITY.
FT   DISULFID   649    663       BY SIMILARITY.
FT   DISULFID   665    674       BY SIMILARITY.
FT   DISULFID   681    692       BY SIMILARITY.
FT   DISULFID   686    701       BY SIMILARITY.
FT   DISULFID   703    712       BY SIMILARITY.
FT   DISULFID   719    729       BY SIMILARITY.
FT   DISULFID   724    738       BY SIMILARITY.
FT   DISULFID   740    749       BY SIMILARITY.
FT   DISULFID   756    767       BY SIMILARITY.
FT   DISULFID   761    776       BY SIMILARITY.
FT   DISULFID   778    787       BY SIMILARITY.
FT   DISULFID   794    805       BY SIMILARITY.
FT   DISULFID   799    814       BY SIMILARITY.
FT   DISULFID   816    825       BY SIMILARITY.
FT   DISULFID   832    843       BY SIMILARITY.
FT   DISULFID   837    854       BY SIMILARITY.
FT   DISULFID   856    865       BY SIMILARITY.
FT   DISULFID   872    883       BY SIMILARITY.
FT   DISULFID   877    892       BY SIMILARITY.
FT   DISULFID   894    903       BY SIMILARITY.
FT   DISULFID   910    921       BY SIMILARITY.
FT   DISULFID   915    930       BY SIMILARITY.
FT   DISULFID   932    941       BY SIMILARITY.
FT   DISULFID   986    997       BY SIMILARITY.
FT   DISULFID   991   1006       BY SIMILARITY.
FT   DISULFID  1008   1017       BY SIMILARITY.
FT   DISULFID  1024   1035       BY SIMILARITY.
FT   DISULFID  1029   1044       BY SIMILARITY.
FT   DISULFID  1046   1055       BY SIMILARITY.
FT   DISULFID  1062   1073       BY SIMILARITY.
FT   DISULFID  1067   1082       BY SIMILARITY.
FT   DISULFID  1084   1093       BY SIMILARITY.
FT   DISULFID  1100   1121       BY SIMILARITY.
FT   DISULFID  1115   1130       BY SIMILARITY.
FT   DISULFID  1132   1141       BY SIMILARITY.
FT   DISULFID  1148   1159       BY SIMILARITY.
FT   DISULFID  1153   1168       BY SIMILARITY.
FT   DISULFID  1170   1179       BY SIMILARITY.
FT   DISULFID  1186   1197       BY SIMILARITY.
FT   DISULFID  1191   1206       BY SIMILARITY.
FT   DISULFID  1208   1217       BY SIMILARITY.
FT   DISULFID  1224   1243       BY SIMILARITY.
FT   DISULFID  1237   1252       BY SIMILARITY.
FT   DISULFID  1254   1263       BY SIMILARITY.
FT   DISULFID  1270   1283       BY SIMILARITY.
FT   DISULFID  1275   1292       BY SIMILARITY.
FT   DISULFID  1294   1303       BY SIMILARITY.
FT   DISULFID  1310   1321       BY SIMILARITY.
FT   DISULFID  1315   1333       BY SIMILARITY.
FT   DISULFID  1335   1344       BY SIMILARITY.
FT   DISULFID  1351   1362       BY SIMILARITY.
FT   DISULFID  1356   1371       BY SIMILARITY.
FT   DISULFID  1373   1382       BY SIMILARITY.
FT   DISULFID  1390   1401       BY SIMILARITY.
FT   DISULFID  1395   1412       BY SIMILARITY.
FT   DISULFID  1414   1423       BY SIMILARITY.
FT   CARBOHYD   462    462       POTENTIAL.
FT   CARBOHYD   887    887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          13.9%;  Score 121;  DB 1;  Length 2524;
  Best Local Similarity  45.0%;  Pred. No. 3.52e-08;
  Matches   18;  Conservative    8;  Mismatches   9;  Indels   5;  Gaps   5;

Db   1236 KCFNNGKC-I-DRVGGYNCICPPGFVGERCEGDVNECLSN 1273
          || ::| | | ||   | |:| ||| | :|| : | |:::
Qy      3 KC-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGE 39


RESULT  13
ID   PGBM_HUMAN     STANDARD;      PRT;  4393 AA.
AC   P98160; Q16287;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
DE   PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
GN   HSPG2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92112994.
RA   KALLUNKI P., TRYGGVASON K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. CELL BIOL. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN, AND COLON;
RX   MEDLINE; 92235084.
RA   MURDOCH A.D., DODGE G.R., COHEN I., TUAN R.S., IOZZO R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
```

```
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. BIOL. CHEM. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 1018-1472 FROM N.A.
RC   TISSUE=COLON;
RX   MEDLINE; 91365376.
RA   DODGE G.R., KOVALSZKY I., CHU M.L., HASSELL J.R., MCBRIDE O.W.,
RA   YI H.F., IOZZO R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   GENOMICS 10:673-680(1991).
RN   [4]
RP   SEQUENCE OF 892-1398 FROM N.A.
RC   TISSUE=FIBROSARCOMA;
RX   MEDLINE; 92120660.
RA   KALLUNKI P., EDDY R.L., BYERS M.G., KESTILA M., SHOWS T.B.,
RA   TRYGGVASON K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   GENOMICS 11:389-396(1991).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE; 94052171.
RA   COHEN I.R., GRAESSEL S., MURDOCH A.D., IOZZO R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:10404-10408(1993).
CC   -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC       MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC       CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC       PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC       SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC       SUBSTRATE FOR CELLS.
CC   -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC       DIMERS OR STELLATE STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       22 C2-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62515; G29470; -.
DR   EMBL; M85289; G184427; -.
DR   EMBL; M64283; G184425; -.
DR   EMBL; S76436; G243371; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   MIM; 142461; -.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PFAM; PF00008; EGF; 4.
DR   PFAM; PF00047; ig; 22.
DR   PFAM; PF00052; laminin_B; 3.
DR   PFAM; PF00053; laminin_EGF; 8.
DR   PFAM; PF00054; laminin_G; 3.
DR   PFAM; PF00057; ldl_recept_a; 4.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; BASEMENT MEMBRANE; PROTEOGLYCAN; REPEAT; GLYCOPROTEIN;
KW   HEPARAN SULFATE; LAMININ EGF-LIKE DOMAIN; IMMUNOGLOBULIN FOLD;
KW   EXTRACELLULAR MATRIX; EGF-LIKE DOMAIN.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   4393       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN      22    193       DOMAIN I (UNIQUE, CONTAINS 3 HS SIDE
FT                                CHAINS).
FT   DOMAIN     194    404       DOMAIN II (4 LDLRA REPEATS).
FT   DOMAIN     405    506       DOMAIN II A (1 IGG-REPEAT).
FT   DOMAIN     507   1678       DOMAIN III (SIMILAR TO SHORT ARM OF
FT                                LAMININ A CHAIN).
FT   DOMAIN    1679   3688       DOMAIN IV (SIMILAR TO NEURAL CELL
FT                                ADHESION MOLECULE; 21 IGG REPEATS).
FT   DOMAIN    3689   4393       DOMAIN V (C-TERMINAL G-DOMAIN OF LAMININ
FT                                ALPHA CHAINS AND EGF).
FT   DOMAIN     197    236       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN     283    321       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     323    361       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     366    405       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     405    506       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     523    532       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN     533    732       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN     733    765       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN     766    815       LAMININ EGF-LIKE 2.
FT   DOMAIN     816    873       LAMININ EGF-LIKE 3.
FT   DOMAIN     881    925       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN     926    935       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN     936   1127       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN    1128   1160       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN    1161   1210       LAMININ EGF-LIKE 6.
FT   DOMAIN    1211   1267       LAMININ EGF-LIKE 7.
FT   DOMAIN    1277   1326       LAMININ EGF-LIKE 8.
FT   DOMAIN    1327   1336       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN    1337   1531       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN    1532   1564       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN    1565   1614       LAMININ EGF-LIKE 10.
FT   DOMAIN    1615   1672       LAMININ EGF-LIKE 11.
FT   DOMAIN    1679   1773       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN    1774   1867       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN    1868   1957       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN    1958   2053       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN    2054   2153       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN    2154   2246       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN    2247   2342       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN    2343   2438       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN    2439   2535       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN    2536   2631       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN    2632   2728       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN    2729   2828       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN    2829   2926       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN    2927   3023       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN    3024   3114       IG-LIKE C2-TYPE DOMAIN 16.
FT   DOMAIN    3115   3213       IG-LIKE C2-TYPE DOMAIN 17.
FT   DOMAIN    3214   3300       IG-LIKE C2-TYPE DOMAIN 18.
FT   DOMAIN    3301   3401       IG-LIKE C2-TYPE DOMAIN 19.
FT   DOMAIN    3402   3490       IG-LIKE C2-TYPE DOMAIN 20.
FT   DOMAIN    3491   3576       IG-LIKE C2-TYPE DOMAIN 21.
FT   DOMAIN    3577   3671       IG-LIKE C2-TYPE DOMAIN 22.
FT   DOMAIN    3701   3847       LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT   DOMAIN    3846   3883       EGF-LIKE 1.
FT   DOMAIN    3886   3924       EGF-LIKE 2.
FT   DOMAIN    3966   4104       LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT   DOMAIN    4106   4143       EGF-LIKE 3.
FT   DOMAIN    4145   4178       EGF-LIKE 4.
FT   DOMAIN    4243   4391       LAMININ G-LIKE 3 (GLOBULAR DOMAIN V C).
FT   SITE        65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE        71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE        76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE      4151   4153       MEDIATES MOTOR NEURON ATTACHMENT
```

```
FT                                     (POTENTIAL).
FT   SITE       4301   4303        MEDIATES MOTOR NEURON ATTACHMENT
FT                                     (POTENTIAL).
FT   DISULFID    199    212        BY SIMILARITY.
FT   DISULFID    206    225        BY SIMILARITY.
FT   DISULFID    219    234        BY SIMILARITY.
FT   DISULFID    285    297        BY SIMILARITY.
FT   DISULFID    292    310        BY SIMILARITY.
FT   DISULFID    304    319        BY SIMILARITY.
FT   DISULFID    325    337        BY SIMILARITY.
FT   DISULFID    332    350        BY SIMILARITY.
FT   DISULFID    344    359        BY SIMILARITY.
FT   DISULFID    368    381        BY SIMILARITY.
FT   DISULFID    375    394        BY SIMILARITY.
FT   DISULFID    388    403        BY SIMILARITY.
FT   DISULFID    766    775        BY SIMILARITY.
FT   DISULFID    768    782        BY SIMILARITY.
FT   DISULFID    785    794        BY SIMILARITY.
FT   DISULFID    797    813        BY SIMILARITY.
FT   DISULFID    816    831        BY SIMILARITY.
FT   DISULFID    818    841        BY SIMILARITY.
FT   DISULFID    844    853        BY SIMILARITY.
FT   DISULFID    856    871        BY SIMILARITY.
FT   DISULFID   1161   1170        BY SIMILARITY.
FT   DISULFID   1163   1177        BY SIMILARITY.
FT   DISULFID   1180   1189        BY SIMILARITY.
FT   DISULFID   1192   1208        BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          13.5%;  Score 118;  DB 1;  Length 4393;
  Best Local Similarity  45.2%;  Pred. No. 1.37e-07;
  Matches   14;  Conservative   10;  Mismatches   4;  Indels   3;  Gaps   3;

Db   3855 CQNGGQCHDSESS-SYVCVCPAGFTGSRCEH 3884
          |::| ||| |: : :| |:| :||:| :||:
Qy      4 CHHG-QCHISDRGEPY-CLCQPGFSGHHCEQ 32


RESULT   14
ID   LI12_CAEEL     STANDARD;      PRT;  1429 AA.
AC   P14585;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   LIN-12 PROTEIN PRECURSOR.
GN   LIN-12 OR R107.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 88334747.
RA   YOCHEM J., WESTON K., GREENWALD I.;
RT   "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
RT   protein with overall similarity to Drosophila Notch.";
RL   NATURE 335:547-550(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
CC   -!- FUNCTION: LIN-12 IS IS INVOLVED IN SEVERAL CELL FATES DECISIONS
CC       THAT REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
CC       ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
CC       EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
CC   -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M12069; G156358; -.
DR   EMBL; Z14092; E1348691; -.
DR   PIR; S06434; S06434.
DR   WORMPEP; R107.8; CE00274.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 12.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 13.
DR   PFAM; PF00023; ank; 4.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16   1429       LIN-12 PROTEIN.
FT   DOMAIN       16    908       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    909    931       POTENTIAL.
FT   DOMAIN      932   1429       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24    618       13 X EGF-TYPE REPEATS.
FT   DOMAIN      631    750       3 X LIN/NOTCH REPEATS.
FT   DOMAIN     1046   1266       6 X ANK MOTIF REPEATS.
FT   DOMAIN       20     61       EGF-LIKE 1.
FT   DOMAIN      114    150       EGF-LIKE 2.
FT   DOMAIN      152    190       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      201    246       EGF-LIKE 4.
FT   DOMAIN      250    285       EGF-LIKE 5.
FT   DOMAIN      287    323       EGF-LIKE 6.
FT   DOMAIN      323    363       EGF-LIKE 7.
FT   DOMAIN      365    402       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      404    441       EGF-LIKE 9.
FT   DOMAIN      449    492       EGF-LIKE 10.
FT   DOMAIN      503    541       EGF-LIKE 11.
FT   DOMAIN      543    579       EGF-LIKE 12.
FT   DOMAIN      582    619       EGF-LIKE 13.
FT   REPEAT      635    669       LIN/NOTCH 1.
FT   REPEAT      670    710       LIN/NOTCH 2.
FT   REPEAT      711    750       LIN/NOTCH 3.
FT   REPEAT     1046   1078       ANK MOTIF 1.
FT   REPEAT     1079   1119       ANK MOTIF 2.
FT   REPEAT     1120   1152       ANK MOTIF 3.
FT   REPEAT     1153   1188       ANK MOTIF 4.
FT   REPEAT     1189   1232       ANK MOTIF 5.
FT   REPEAT     1233   1266       ANK MOTIF 6.
FT   DISULFID     24     35       BY SIMILARITY.
FT   DISULFID     29     49       BY SIMILARITY.
FT   DISULFID     51     60       BY SIMILARITY.
FT   DISULFID    118    129       BY SIMILARITY.
FT   DISULFID    123    138       BY SIMILARITY.
FT   DISULFID    140    149       BY SIMILARITY.
FT   DISULFID    156    169       BY SIMILARITY.
FT   DISULFID    163    178       BY SIMILARITY.
```

```
FT   DISULFID   180   189       BY SIMILARITY.
FT   DISULFID   205   227       BY SIMILARITY.
FT   DISULFID   221   234       BY SIMILARITY.
FT   DISULFID   236   245       BY SIMILARITY.
FT   DISULFID   254   264       BY SIMILARITY.
FT   DISULFID   259   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   302       BY SIMILARITY.
FT   DISULFID   296   311       BY SIMILARITY.
FT   DISULFID   313   322       BY SIMILARITY.
FT   DISULFID   327   339       BY SIMILARITY.
FT   DISULFID   334   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   429       BY SIMILARITY.
FT   DISULFID   431   440       BY SIMILARITY.
FT   DISULFID   507   518       BY SIMILARITY.
FT   DISULFID   512   529       BY SIMILARITY.
FT   DISULFID   531   540       BY SIMILARITY.
FT   DISULFID   547   558       BY SIMILARITY.
FT   DISULFID   552   567       BY SIMILARITY.
FT   DISULFID   569   578       BY SIMILARITY.
FT   DISULFID   586   597       BY SIMILARITY.
FT   DISULFID   591   607       BY SIMILARITY.
FT   DISULFID   609   618       BY SIMILARITY.
FT   CARBOHYD    41    41       POTENTIAL.
FT   CARBOHYD   165   165       POTENTIAL.
FT   CARBOHYD   194   194       POTENTIAL.
FT   CARBOHYD   378   378       POTENTIAL.
FT   CARBOHYD   515   515       POTENTIAL.
FT   CARBOHYD   623   623       POTENTIAL.
FT   CARBOHYD   751   751       POTENTIAL.
FT   CARBOHYD   754   754       POTENTIAL.
FT   CARBOHYD   900   900       POTENTIAL.
SQ   SEQUENCE   1429 AA;  157115 MW;  CFD2CCA4 CRC32;

  Query Match          13.2%;  Score 115;  DB 1;  Length 1429;
  Best Local Similarity  43.6%;  Pred. No. 5.24e-07;
  Matches   17;  Conservative    9;  Mismatches  10;  Indels   3;  Gaps   3;

Db    334 CNHGTCIDSPLSEKAFECQCEPGYEGILCEQDKNECLSE 372
          |:|| |  |  :| :: | |:||: |  |||: | |::|
Qy      4 CHHGQCHISDRGE-PY-CLCQPGFSGHHCEQE-NPCMGE 39

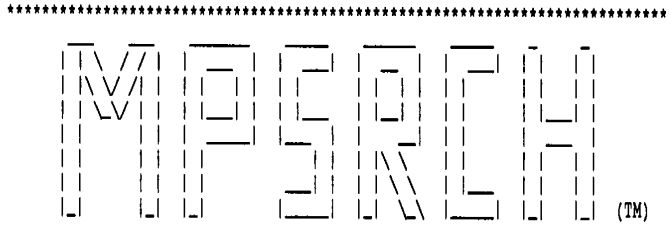

RESULT   15
     CYR6_HUMAN     STANDARD;      PRT;   381 AA.
     000622; O14934;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (GIG1 PROTEIN) (INSULIN-LIKE GROWTH FACTOR-
DE   BINDING PROTEIN 10).
GN   IGFBP10 OR CYR61 OR GIG1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALBRECHT C., VON DER KAMMER H., KLAUDINY J., MAYHAUS M., NITSCH R.M.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97280750.
RA   JAY P., BERGE-LEFRANC J.L., MARSOLLIER C., MEJEAN C., TAVIAUX S.,
RA   BERTA P.;
RT   "The human growth factor-inducible immediate early gene, CYR61, maps
RT   to chromosome 1p.";
RL   ONCOGENE 14:1753-1757(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   KOLESNIKOVA T.V., LAU L.F.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   BI A.B., YU L.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12084; E311857; -.
DR   EMBL; U62015; G2130527; -.
DR   EMBL; AF003594; G2196782; -.
DR   EMBL; AF031385; G2606094; -.
DR   MIM; 602369; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25   381       CYR61 PROTEIN.
FT   DOMAIN      98   164       VWFC.
FT   DOMAIN     286   360       CTCK.
FT   DISULFID   286   323       BY SIMILARITY.
FT   DISULFID   303   337       BY SIMILARITY.
FT   DISULFID   314   353       BY SIMILARITY.
FT   DISULFID   317   355       BY SIMILARITY.
FT   DISULFID   322   359       BY SIMILARITY.
FT   CONFLICT   210   210       L -> I (IN REF. 4).
FT   CONFLICT   220   220       L -> R (IN REF. 4).
SQ   SEQUENCE   381 AA;  42026 MW;  2B091D9E CRC32;

  Query Match          13.1%;  Score 114;  DB 1;  Length 381;
  Best Local Similarity  32.3%;  Pred. No. 8.18e-07;
  Matches   20;  Conservative   10;  Mismatches  29;  Indels   3;  Gaps   3;

Db    300 YAGCLSVKKYRPKYC-GSCVDGRCCTPQLTRTVKMRFRCEDGETFSKNVMMIQSCKCNYN 358
          ||:| :  |   : | |:|    || |  ::  |  |:| || :|  :|     | |
Qy     51 YASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCR-A 108

Db    359 CP 360
          |:
Qy    109 CS 110


Search completed: Fri May 28 09:25:47 1999
Job time : 19 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

un on: Fri May 28 09:26:05 1999; MasPar time 10.25 Seconds
585.992 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-11
Description: (1-110) from US09191647.pep
Perfect Score: 873
Sequence: 1 AFKCHHGQCHISDRGEPYCL.....GSSFVEEVERHLECGCRACS 110

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 37.119; Variance 57.890; scale 0.641

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	836	95.8	1523	11	O88280	MEGF5.	3.63e-188
2	798	91.4	739	4	O75094	MEGF5 (FRAGMENT).	4.81e-178
3	442	50.6	1531	11	O88279	MEGF4.	9.62e-85
4	310	35.5	79	4	O75093	MEGF4 (FRAGMENT).	1.76e-51
5	185	21.2	601	5	Q20204	F40E10.4 PROTEIN (FRAG	9.57e-22
6	128	14.7	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	1.79e-09
7	124	14.2	2447	13	O13149	NOTCH 2 (FRAGMENT).	1.14e-08
8	123	14.1	434	11	O55139	JAGGED2 PROTEIN (FRAGM	1.81e-08
9	123	14.1	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	1.81e-08
10	123	14.1	1202	11	P97607	JAGGED2 (FRAGMENT).	1.81e-08
11	121	13.9	955	4	Q99466	NOTCH4 (FRAGMENT).	4.52e-08
12	121	13.9	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	4.52e-08
13	121	13.9	1964	11	O35442	NOTCH4.	4.52e-08
14	121	13.9	1999	4	Q99940	NOTCH4.	4.52e-08
15	121	13.9	2003	4	O00306	NOTCH4.	4.52e-08
16	121	13.9	2470	11	O35516	CELL SURFACE PROTEIN.	4.52e-08
17	118	13.5	169	4	O14944	EPIREGULIN.	1.76e-07
18	117	13.4	530	5	Q24526	SLIT LOCUS ENCODING A	2.77e-07
19	117	13.4	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	2.77e-07
20	116	13.3	762	13	O42373	NOTCH RECEPTOR PROTEIN	4.33e-07
21	115	13.2	372	5	Q21756	HYPOTHETICAL 39.1 KD P	6.78e-07
22	115	13.2	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	6.78e-07
23	115	13.2	2352	5	O61240	HRNOTCH PROTEIN.	6.78e-07
24	114	13.1	381	4	O43775	CYR61 PROTEIN.	1.06e-06
25	114	13.1	728	13	Q90656	TRANSMEMBRANE PROTEIN	1.06e-06
26	112	12.8	367	11	O54775	ELM1.	2.57e-06
27	111	12.7	752	13	O42374	NOTCH RECEPTOR PROTEIN	3.99e-06
28	110	12.6	717	13	P87357	DELTAD TRANSMEMBRANE P	6.19e-06
29	109	12.5	162	11	Q61521	EPIREGULIN.	9.59e-06
30	109	12.5	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	9.59e-06
31	109	12.5	2180	5	O01768	SIMILARITY TO EGF-LIKE	9.59e-06
32	108	12.4	308	6	O46370	PREADIPOCYTE FACTOR-1.	1.48e-05
33	108	12.4	406	5	Q25059	FIBROPELLIN III (FRAGM	1.48e-05
34	108	12.4	1095	4	Q99458	NOTCH4 (FRAGMENT).	1.48e-05
35	108	12.4	1476	13	Q90285	PUTATIVE EXTRACELLULAR	1.48e-05
36	107	12.3	383	11	O70534	ZOG.	2.29e-05
37	107	12.3	383	11	Q62779	PREADIPOCYTE FACTOR 1.	2.29e-05
38	107	12.3	502	5	O17692	MEC-9 PROTEIN.	2.29e-05
39	107	12.3	661	5	O61537	SPERM TRANSMEMBRANE PR	2.29e-05
40	107	12.3	838	5	Q18761	MEC-9 PROTEIN.	2.29e-05
41	107	12.3	838	5	Q27422	MEC-9L.	2.29e-05
42	106	12.1	293	5	Q20979	F58E6.3 PROTEIN.	3.52e-05
43	106	12.1	387	11	Q06007	NOTCH PROTEIN HOMOLOG	3.52e-05
44	106	12.1	802	13	O57462	DELTAA.	3.52e-05
45	106	12.1	3857	11	O88840	MUTANT FIBRILLIN-1.	3.52e-05

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 95.8%; Score 836; DB 11; Length 1523;
Best Local Similarity 94.5%; Pred. No. 3.63e-188;
Matches 104; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1414 AFKCHHGQCHISDRGEPYCLCQPGFSGNHCEQENPCLGEIVREAIRRQKDYASCATASKV 1473
||||||||||||||||||||||||||:||||||||:||||||||||||||||||||||
Qy 1 AFKCHHGQCHISDRGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASKV 60

Db 1474 PIMVCRGGCGSQCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRECS 1523
||| |||||: ||||||||||||||||||||||||||||||||||| ||
Qy 61 PIMECRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRACS 110

RESULT 2
ID O75094 PRELIMINARY; PRT; 739 AA.
AC O75094;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5 (FRAGMENT).
GN MEGF5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011538; D1033429; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1
SQ SEQUENCE 739 AA; 80364 MW; DC6BCB63 CRC32;

Query Match 91.4%; Score 798; DB 4; Length 739;
Best Local Similarity 89.1%; Pred. No. 4.81e-178;
Matches 98; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 630 AFKCHHGQCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKV 689
Qy 1 AFKCHHGQCHISDRGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASKV 60

Db 690 PIMECRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 739
Qy 61 PIMECRGGCGTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRACS 110

RESULT 3
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 50.6%; Score 442; DB 11; Length 1531;
Best Local Similarity 50.0%; Pred. No. 9.62e-85;
Matches 56; Conservative 22; Mismatches 31; Indels 3; Gaps 3;

Db 1421 GLQCLHGHCQASATRG-AHCVCSPGFSGELCEQESECRGDPVRDFHRVQRGYAICQTTRP 1479
Qy 1 AFKCHHGQCHISD-RGEPYCLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATASK 59

Db 1480 LSWVECRGACPGQGCCQGLRLKRRKLTFECSDGTSFAEEVEKPTKCGCAPCA 1531
Qy 60 VPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRACS 110

RESULT 4
ID O75093 PRELIMINARY; PRT; 79 AA.
AC O75093;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4 (FRAGMENT).
GN MEGF4.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011537; D1033428; -.
DR PROSITE; PS01185; CTCK_1; 1.
FT NON_TER 1 1
SQ SEQUENCE 79 AA; 8809 MW; 96C95FFE CRC32;

Query Match 35.5%; Score 310; DB 4; Length 79;
Best Local Similarity 48.1%; Pred. No. 1.76e-51;
Matches 38; Conservative 17; Mismatches 23; Indels 1; Gaps 1;

Db 1 ESECRGDPVRDFHQVQRGYAICQTTRPLSWVECRGSCPGQGCCQGLRLKRRKFTFECSDG 60
Qy 33 ENPCMGEIVREAIRRQKDYASCATASKVPIMECRGGC-GTTCCQPIRSKRRKYVFQCTDG 91

Db 61 TSFAEEVEKPTKCGCALCA 79
Qy 92 SSFVEEVERHLECGCRACS 110

RESULT 5
ID Q20204 PRELIMINARY; PRT; 601 AA.
AC Q20204;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F40E10.4 PROTEIN (FRAGMENT).
GN F40E10.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA SMYE R.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z69792; E1346469; -.
DR PROSITE; PS01187; EGF_CA; 1.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 1 1

```
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match          21.2%;  Score 185;  DB 5;  Length 601;
  Best Local Similarity  28.2%;  Pred. No. 9.57e-22;
  Matches   31;  Conservative   26;  Mismatches  46;  Indels   7;  Gaps   5;

Db    485 GIDCGNGKCTNNALSPKGYMCQCDSHFSGEHCD-EKRIKCD--KQKFRRHHIENECRSVD 541
          :: | :| |  :      | | |:: ||| ||: |:  : :  :: :||:     | :
Qy      1 AFKCHHGQCHISD-RGEPY-CLCQPGFSGHHCEQENPCMGEIVREAIRRQKDYASCATAS 58

Db    542 RIKIAECNGYCGGEQNCCTAVKKKQRKVKMICKNGTTKISTVHIIRQCQC 591
          :: | || | ||    || ::: |:||    | :|:: :  |    :| |
Qy     59 KVPIMECRGGCGT--TCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGC 106


RESULT   6
ID   Q19350      PRELIMINARY;      PRT;  1722 AA.
AC   Q19350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO EGF-LIKE REPEATS. NCBI GI: 1125776.
GN   F11C7.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A., VETTER J.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U42839; G1125776; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PFAM; PF00008; EGF; 24.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1722 AA;  188383 MW;  CCFB86B8 CRC32;

  Query Match          14.7%;  Score 128;  DB 5;  Length 1722;
  Best Local Similarity  38.9%;  Pred. No. 1.79e-09;
  Matches   14;  Conservative    8;  Mismatches  13;  Indels   1;  Gaps   1;

Db    372 IRCLNGGSCKLDAEGEPFCVCEEGFDGPFCEPKSGC 407
          ::| :|  | :   |||:|:|: || |  ||  : |
Qy      2 FKCHHG-QCHISDRGEPYCLCQPGFSGHHCEQENPC 36


RESULT   7
ID   O13149      PRELIMINARY;      PRT;  2447 AA.
AC   O13149;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH 2 (FRAGMENT).
OS   FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   TETRAODONTIFORMES; TETRAODONTOIDEI; TETRAODONTIDAE; FUGU.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NAKAMURA T., TROWSDALE J.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB004829; D1021371; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE   2447 AA;  262542 MW;  3CDA4F7A CRC32;

  Query Match          14.2%;  Score 124;  DB 13;  Length 2447;
  Best Local Similarity  42.9%;  Pred. No. 1.14e-08;
  Matches   15;  Conservative    8;  Mismatches  10;  Indels   2;  Gaps   2;

Db    683 CVHGKC-IEQQNGYFCQCEAGWVGQHCEQEKDECL 716
          | || | | ::   :| |::|  |:|||| : |:
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCM 37


RESULT   8
ID   O55139      PRELIMINARY;      PRT;  434 AA.
AC   O55139;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   JAGGED2 PROTEIN (FRAGMENT).
GN   JAG2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   VALSECCHI V., BALLABIO A., RUGARLI E.I.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Y14495; E1227811; -.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 5.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
FT   NON_TER    434    434
SQ   SEQUENCE   434 AA;  46176 MW;  D68AE029 CRC32;

  Query Match          14.1%;  Score 123;  DB 11;  Length 434;
  Best Local Similarity  36.4%;  Pred. No. 1.81e-08;
  Matches   16;  Conservative   13;  Mismatches  11;  Indels   4;  Gaps   4;

Db    282 CGPHGHC-VSLPGGNFSCICDSGFTGTYCHENIDDCMGQPCRNG 324
          |  ||:| :|  |  : |:|::||:|  | :: : |||:  |::
Qy      4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44


RESULT   9
ID   O70219      PRELIMINARY;      PRT;  518 AA.
AC   O70219;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   JAGGED 2 (JAGGED 2 PROTEIN) (FRAGMENT).
GN   JAG2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=BRAIN;
RA   LAN Y., JIANG R., SHAWBER C., WEINMASTER G., GRIDLEY T.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF010137; G3057059; -.
DR   MGD; MGI:1098270; JAG2.
DR   PROSITE; PS01187; EGF_CA; 5.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     518    518
SQ   SEQUENCE   518 AA;  55119 MW;  7144EB20 CRC32;

  Query Match           14.1%;  Score 123;  DB 11;  Length 518;
  Best Local Similarity  36.4%;  Pred. No. 1.81e-08;
  Matches    16;  Conservative   13;  Mismatches  11;  Indels   4;  Gaps   4;

Db    305 CGPHGHC-VSLPGGNFSCICDSGFTGTYCHENIDDCMGQPCRNG 347
          |  ||:| :|  |  : |:|::||:|  | :: : |||:  |::
Qy      4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44


RESULT   10
[redacted]  P97607      PRELIMINARY;      PRT;  1202 AA.
[redacted]  P97607;
[redacted]  01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   JAGGED2 (FRAGMENT).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97105852.
RA   SHAWBER C.J., BOULTER J., LINDSELL C.E., WEINMASTER G.;
RT   "Jagged2: a serrate-like gene expressed during rat embryogenesis.";
RL   DEV. BIOL. 180:370-376(1996).
DR   EMBL; U70050; G1718248; -.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   1202 AA;  129703 MW;  697F4205 CRC32;

  Query Match           14.1%;  Score 123;  DB 11;  Length 1202;
  Best Local Similarity  36.4%;  Pred. No. 1.81e-08;
  Matches    16;  Conservative   13;  Mismatches  11;  Indels   4;  Gaps   4;

Db    561 CGPHGHC-VSLPGGNFSCICDSGFTGTYCHENIDDCMGQPCRNG 603
          |  ||:| :|  |  : |:|::||:|  | :: : |||:  |::
[redacted]    4 C-HHGQCHISDRGEPY-CLCQPGFSGHHCEQE-NPCMGEIVREA 44


RESULT   11
ID   Q99466      PRELIMINARY;      PRT;   955 AA.
AC   Q99466;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4 (FRAGMENT).
GN   NOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97311416.
RA   SUGAYA K., SASANUMA S., NOHATA J., KIMURA T., FUKAGAWA T.,
RA   NAKAMURA Y., ANDO A., INOKO H., IKEMURA T., MITA K.;
RT   "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
RT   human counterpart gene of mouse proto-oncogene Int3.";
RL   GENE 189:235-244(1997).
DR   EMBL; D86566; D1013803; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 21.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     955    955
SQ   SEQUENCE   955 AA;  100017 MW;  28507B36 CRC32;

  Query Match           13.9%;  Score 121;  DB 4;  Length 955;
  Best Local Similarity  43.2%;  Pred. No. 4.52e-08;
  Matches    16;  Conservative    6;  Mismatches  12;  Indels   3;  Gaps   3;

Db    401 CHGDAQCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 437
          ||  :||  :   |   ||||||:||  | |: : |:
Qy      4 CH-HGQCHISD-RGEPYCLCQPGFSGHHCEQE-NPCM 37


RESULT   12
ID   Q61204      PRELIMINARY;      PRT;  1687 AA.
AC   Q61204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH2-LIKE (EGF REPEAT TRANSMEMBRANE PROTEIN).
GN   NOTCH2L.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=WHOLE EMBRYO;
RA   SELL C., HOFF III H.B.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U57368; G1336628; -.
DR   MGD; MGI:1202397; NOTCH2L.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PFAM; PF00008; EGF; 7.
KW   TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1687 AA;  188528 MW;  73B9DDDC CRC32;

  Query Match           13.9%;  Score 121;  DB 11;  Length 1687;
  Best Local Similarity  37.8%;  Pred. No. 4.52e-08;
  Matches    14;  Conservative   11;  Mismatches   9;  Indels   3;  Gaps   3;

Db    344 CQNGGTCHMLSR-DTYECTCQVGFTGKQCQWTDACLS 379
          |::|  ||:  | : | | || ||:| :|:  ::|::
Qy      4 CHHG-QCHISDRGEPY-CLCQPGFSGHHCEQENPCMG 38


RESULT   13
ID   O35442      PRELIMINARY;      PRT;  1964 AA.
AC   O35442;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   NOTCH4.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROWEN L., MAHAIRAS G., QIN S., AHEARN M.E., DANKERS C., LASKY S.,
RA   LORETZ C., SCHMIDT S., TIPTON S., TRAICOFF R., ZACKRONE K., HOOD L.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF030001; G2564947; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
```

```
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1964 AA;  206699 MW;  CE2CA3B6 CRC32;

  Query Match          13.9%;  Score 121;  DB 11;  Length 1964;
  Best Local Similarity  40.5%;  Pred. No. 4.52e-08;
  Matches   15;  Conservative    8;  Mismatches  13;  Indels   1;  Gaps   1;

Db    443 CEHGGSCINTPGSFNCLCLPGYTGSRCEADHNECLSQ 479
          | ||   |:  |   ||| ||::| :|| : | |:::
Qy      4 CHHGQCHISDRGEPYCLCQPGFSGHHCEQE-NPCMGE 39


RESULT   14
ID   Q99940      PRELIMINARY;      PRT;  1999 AA.
AC   Q99940;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   NOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LI L., HUANG G., BANTA A., DENG Y., CHEN L., PHAM Q., ROWEN L.,
RA   HOOD L.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U89335; G1841543; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 5.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1999 AA;  209134 MW;  0680278E CRC32;

  Query Match          13.9%;  Score 121;  DB 4;  Length 1999;
  Best Local Similarity  43.2%;  Pred. No. 4.52e-08;
  Matches   16;  Conservative    6;  Mismatches  12;  Indels   3;  Gaps   3;

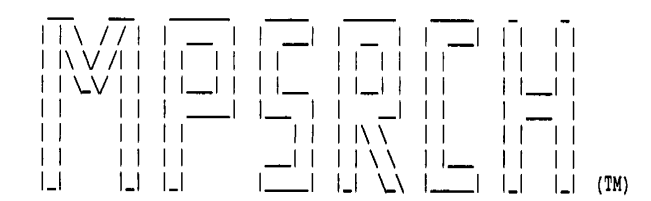
Db    400 CHGDAQCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 436
          ||  :||  :   |   ||||||:||  | |: : |:
Qy      4 CH-HGQCHISD-RGEPYCLCQPGFSGHHCEQE-NPCM 37


RESULT   15
ID   O00306      PRELIMINARY;      PRT;  2003 AA.
AC   O00306;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH4.
GN   HNOTCH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW, HEART;
RA   LI L., HUANG G., BANTA A., YU D., ROWEN L., HOOD L.;
RL   SUBMITTED (MAR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U95299; G2072309; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 5.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2003 AA;  209620 MW;  518CFE96 CRC32;

  Query Match          13.9%;  Score 121;  DB 4;  Length 2003;
  Best Local Similarity  43.2%;  Pred. No. 4.52e-08;
  Matches   16;  Conservative    6;  Mismatches  12;  Indels   3;  Gaps   3;

Db    401 CHGDAQCSTNPLTGSTLCLCQPGYSGPTCHQDLDECL 437
          ||  :||  :   |   ||||||:||  | |: : |:
Qy      4 CH-HGQCHISD-RGEPYCLCQPGFSGHHCEQE-NPCM 37


Search completed: Fri May 28 09:26:46 1999
Job time : 41 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:28:42 1999; MasPar time 7.62 Seconds
374.148 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-12
Description: (1-134) from US09191647.pep
Perfect Score: 971
Sequence: 1 HLRVLQLMENRISTIERGAF.....SFNHMPKLRTFRLHSNNLYC 134

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 30.391; Variance 166.888; scale 0.182

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	779	80.2	1534	30	W46966	Amino acid sequence o	4.48e-49
2	357	36.8	1480	5	R25079	Drosophila SLIT prote	3.21e-17
3	267	27.5	1091	27	W41641	Sequence used in dete	1.16e-10
4	258	26.6	560	12	R71294	Human glycoprotein V.	5.12e-10
5	252	26.0	605	17	R85888	WD-40 domain-contg. i	1.38e-09
6	242	24.9	603	17	R85889	WD-40 domain-contg. r	7.10e-09
7	220	22.7	353	1	R05160	Sequence of human bon	2.56e-07
8	215	22.1	369	15	R87952	Human neurotrophic bi	5.75e-07
9	213	21.9	368	1	R05159	Sequence of human bon	7.94e-07
10	213	21.9	369	15	R87951	Rat neurotrophic bigl	7.94e-07
11	210	21.6	332	15	R87953	Bovine neurotrophic b	1.29e-06
12	196	20.2	234	8	R42265	Decorin sequence PT-7	1.22e-05
13	196	20.2	280	8	R42266	Decorin sequence PT-7	1.22e-05
14	196	20.2	305	8	R42267	Decorin sequence PT-7	1.22e-05
15	196	20.2	331	8	R42260	Mature decorin PT-65.	1.22e-05
16	196	20.2	342	17	R89439	Human recombinant dec	1.22e-05
17	196	20.2	345	23	W09405	Pineal gland specific	1.22e-05
18	196	20.2	1388	18	R89471	Collagen/decorin fusi	1.22e-05
19	194	20.0	186	8	R42264	Decorin sequence PT-7	1.67e-05
20	194	20.0	904	39	W86351	Human DNAX toll-like	1.67e-05
21	194	20.0	1112	16	R85298	Tomato pathogen resis	1.67e-05
22	194	20.0	1112	16	R85299	Tomato pathogen resis	1.67e-05
23	192	19.8	196	5	R29102	Drosophila SLIT prote	2.30e-05
24	191	19.7	610	9	R51116	Platelet glycoprotein	2.70e-05
25	191	19.7	610	9	R56664	Mutant platelet glyco	2.70e-05
26	191	19.7	610	23	W18201	Platelet glycoprotein	2.70e-05
27	191	19.7	610	17	R89436	Mutated platelet glyc	2.70e-05
28	184	18.9	976	21	W13408	Arabidopsis thaliana	8.20e-05
29	180	18.5	293	1	P91368	45 kDa amino terminal	1.54e-04
30	177	18.2	799	39	W86352	Human DNAX toll-like	2.48e-04
31	177	18.2	837	39	W86361	Human DNAX toll-like	2.48e-04
32	172	17.7	376	7	R36773	Human fibromodulin.	5.43e-04
33	172	17.7	376	24	W26404	Human fibromodulin.	5.43e-04
34	162	16.7	614	32	W58382	Human secreted protei	2.58e-03
35	159	16.4	784	29	W48245	Human pro-tumour necr	4.11e-03
36	159	16.4	784	39	W86350	Human DNAX toll-like	4.11e-03
37	159	16.4	784	39	W90069	Human TNF-alpha conve	4.11e-03
38	158	16.3	1874	35	W64518	Adenylate cyclase pro	4.79e-03
39	157	16.2	394	39	W86363	Mouse DNAX toll-like	5.59e-03
40	156	16.1	139	8	R42263	Decorin sequence PT-7	6.53e-03
41	156	16.1	1045	39	W86354	Human DNAX toll-like	6.53e-03
42	154	15.9	644	39	W82318	Human 7-transmembrane	8.88e-03
43	150	15.4	764	5	R24244	Rat thyrotropin recep	1.64e-02
44	149	15.3	342	6	R30492	N-terminal of LH rece	1.91e-02
45	148	15.2	762	21	W14778	Human TSH receptor.	2.23e-02

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 80.2%; Score 779; DB 30; Length 1534;
Best Local Similarity 77.6%; Pred. No. 4.48e-49;
Matches 104; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

```
Db    86 qlrvlqlmenqigavergafddmkelerlrlnrnqlhmlpellfqnnqalsrldlsenai 145
         :|||||||||:|:::|||||:|:||||||||||| |:::|||||     | ||||||| |
Qy     1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db   146 qaiprkafrgatdlknlrldknqiscieegafralrglevltlnnnnittipvssfnhmp 205
         |||||||||| |:|||:|| |||||||:||||||| |||||||||||| ::|:||||||
```

```
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    206 klrtfrlhsnhlfc 219
          ||||||||||:|:|
Qy    121 KLRTFRLHSNNLYC 134


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;

  Query Match          36.8%;  Score 357;  DB 5;  Length 1480;
  Best Local Similarity  39.2%;  Pred. No. 3.21e-17;
  Matches   51;  Conservative   32;  Mismatches  47;  Indels   0;  Gaps   0;

Db   105 lelqgnnltviyetdfqrltklrmlqltdnqihtiernsfqdlvslerldisnnvittvg 164
          |:|  | :: |    || |  |  |:|  | :: :    |     | |||:|:| | ::
Qy     5 LQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQAIP 64

Db   165 rrvfkgaqslrslqldnnqitcldehafkglveleiltvnnnnltslphnifggvgrlra 224
          |: |:||  :::|||| |||:|::: ||::| :||:||:||||:| |:   |  : :||:
Qy    65 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 124

Db   225 lrlsdnpfac 234
          :||  | : |
Qy   125 FRLHSNNLYC 134


RESULT    3
ID   W41641 standard; Protein; 1091 AA.
AC   W41641;
DT   27-APR-1998  (first entry)
DE   Sequence used in detection method.
KW   Detection; mouse; murine.
OS   Mus sp.
PN   J09107971-A.
PD   28-APR-1997.
PF   19-OCT-1995; 270822.
PR   19-OCT-1995; JP-270822.
PA   (TANA ) TANABE SEIYAKU CO.
PA   (TOYA/) TOYAMA S.
DR   WPI; 97-292464/27.
DR   N-PSDB; V04445.
PT   Detection of genes, useful for cloning genes of high and low
PT   expression - by homogenising prepared ds-cDNA pool from sample for
PT   comparison with each other to remove specific DNA fragment
PS   Claim 9; Pages 13-16; 18pp; Japanese.
CC   The present sequence was used in the development of a novel method
CC   for detecting genes showing difference in expression quantity
CC   between several samples. The method comprises preparing a double
CC   stranded cDNA pool of a standard organism sample, and homogenising
CC   the contents of each DNA fragment contained in the resultant cDNA
CC   pool to prepare a content homogenised standard cDNA pool. Double
CC   stranded cDNA pools derived from each sample for several organism
CC   samples are compared with each other, and the DNA fragment
CC   associated with the DNA fragment in the cDNA pool from the content
CC   homogenised standard cDNA pool is removed to give a remaining cDNA
CC   pool for each of the samples to be compared. The method can clone
CC   gene groups of high and low expression level efficiently.
SQ   Sequence   1091 AA;

  Query Match          27.5%;  Score 267;  DB 27;  Length 1091;
  Best Local Similarity  35.3%;  Pred. No. 1.16e-10;
  Matches   47;  Conservative   37;  Mismatches  47;  Indels   2;  Gaps   2;

Db   166 rirelnlasnrisilesgafdglsrslltlrlsknritqlpvkaf-klprltqldlnrnr 224
          ::| | |  |||| :| |||: | : |  |||::| :  :|   |  :|| :|||: |:
Qy     1 HLRVLQLMENRISTIERGAFQDL-KELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db   225 irliegltfqgldslevlrlqrnnisrltdgafwglskmhvlhleynslvevnsgslygl 284
          |: |   :|:|   :  |:|: | || : ||||::|  : || |: |::  :: :|:  :
Qy    60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db   285 talhqlhlsnnsi 297
            |: ::| :|::
Qy   120 PKLRTFRLHSNNL 132


RESULT    4
```

ID R71294 standard; Protein; 560 AA.
AC R71294;
DT 18-AUG-1995 (first entry)
DE Human glycoprotein V.
KW Glycoprotein V; GPV; platelet.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..16
FT /label= Sig_peptide
FT modified_site 51
FT /label= N-glycosylation_site
FT modified_site 181
FT /label= N-glycosylation_site
FT modified_site 244
FT /label= N-glycosylation_site
FT modified_site 267
FT /note= "N-glycosylation_site"
FT modified_site 298
FT /label= N-glycosylation_site
FT modified_site 312
FT /label= N-glycosylation_site
FT modified_site 385
FT /label= N-glycosylation_site
FT cleavage_site 476..477
FT /note= "putative thrombin cleavage site"
FT modified_site 499
FT /label= N-glycosylation_site
FT domain 520..544
FT /note= "putative transmembrane domain"
PN WO9502054-A.
PD 19-JAN-1995.
PF 07-JUL-1994; U07644.
PR 09-JUL-1993; US-089455.
PR 03-DEC-1993; US-162599.
PR 10-FEB-1994; US-195006.
PA (CORT-) COR THERAPEUTICS INC.
PI Cazenave J, Lanza F, Phillips DR;
DR WPI; 95-066899/09.
DR N-PSDB; Q85594.
PT Platelet glycoprotein V gene - useful for producing glycoprotein
PT V (GPV) and variants and generating antibodies to GPV
PS Disclosure; Page 45-50; 82pp; English.
CC Genomic clones were isolated from a human fibroblast library in
CC lambda Fix using a 748 bp 32P-labeled glycoprotein V (GPV) cDNA
CC probe. Exon-containing fragments from positive clones were
CC subcloned and sequenced. The full sequence of the human GPV gene
CC is given in Q85594.
SQ Sequence 560 AA;

Query Match 26.6%; Score 258; DB 12; Length 560;
Best Local Similarity 32.8%; Pred. No. 5.12e-10;
Matches 43; Conservative 37; Mismatches 51; Indels 0; Gaps 0;

```
Db     76 lqrlmisdshisavapgtfsdliklktlrlsrnkithlpgalldkmvlleqlfldhnalr 135
          |: | : :::||:: |:| || | |||:||:: :| |:    | :| | | ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    136 gidqnmfqklvnlqelalnqnqldflpaslftnlenlklldlsgnnlthlpkgllgaqak 195
          :| :: |: |:: :| |: ||: : : | | :| :| |: ||:|:|: : :   :|
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    196 lerlllhsnrl 206
          | : |||| |
Qy    122 LRTFRLHSNNL 132
```

RESULT 5
ID R85888 standard; Protein; 605 AA.
AC R85888;
DT 13-SEP-1996 (first entry)
DE WD-40 domain-contg. insulin-like growth factor binding protein.
KW WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW intracellular signalling; protein kinase C; homology; motif; modulator;
KW receptors of activated protein kinase; enzyme activity; isozyme; human.
OS Synthetic.
PN WO9521252-A2.
PD 10-AUG-1995.
PF 31-JAN-1995; U01210.
PR 01-FEB-1994; US-190802.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Mochly-Rosen D, Ron D;
DR WPI; 95-283772/37.
PT New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT activity of a protein, eg. protein kinase C, which interacts with a
PT protein contg. a WD-40 region.
PS Example 5; Page 122-125; 351pp; English.
CC Proteins R85851-92 are protein which contain at least one WD-40 (also
CC called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC regions are involved in protein-protein interactions between proteins
CC involved in intracellular signalling. An example of such an interaction
CC is between protein kinase C and receptors of activated protein kinase
CC (RACK), esp. RACK-1 (R85850). Proteins R85851-82 were isolated based on
CC homology with beta-transducin, whereas proteins R85882-92 were isolated
CC based on homology with the WD-40 consensus sequence (R85893). The
CC proteins were used to construct the peptides R84928-R85063 and
CC R85786-R85842. The peptides can be used to identify target proteins
CC contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC proteins involved in protein-protein interaction and to screen for drugs
CC that will affect protein-protein interaction involving WD-40 domains.
SQ Sequence 605 AA;

Query Match 26.0%; Score 252; DB 17; Length 605;
Best Local Similarity 32.8%; Pred. No. 1.38e-09;
Matches 43; Conservative 30; Mismatches 58; Indels 0; Gaps 0;

```
Db    220 lreldlsrnalraikanvfvqlprlqklyldrnliaavapgaflglkalrwldlshnrva 279
          || |:|  | : :|    | :|  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 glledtfpgllglrvlrlshnaiaslrprtfkdlhfleelqlghnrirqlaersfeglgq 339
          ::   :| | : :: |:|  | |: :   :|: |: || | | :| | :|:  ||: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 levltldhnql 350
          |  : |  | |
Qy    122 LRTFRLHSNNL 132
```

RESULT 6
ID R85889 standard; Protein; 603 AA.
AC R85889;
DT 13-SEP-1996 (first entry)
DE WD-40 domain-contg. rat insulin-like growth factor binding protein.
KW WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW intracellular signalling; protein kinase C; homology; motif; modulator;
KW receptors of activated protein kinase; enzyme activity; isozyme; human.
OS Rattus rattus.
PN WO9521252-A2.
PD 10-AUG-1995.
PF 31-JAN-1995; U01210.
PR 01-FEB-1994; US-190802.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Mochly-Rosen D, Ron D;
DR WPI; 95-283772/37.
PT New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT activity of a protein, eg. protein kinase C, which interacts with a
PT protein contg. a WD-40 region.
PS Example 5; Page 125-128; 351pp; English.
CC Proteins R85851-92 are protein which contain at least one WD-40 (also
CC called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC regions are involved in protein-protein interactions between proteins
CC involved in intracellular signalling. An example of such an interaction
CC is between protein kinase C and receptors of activated protein kinase
CC (RACK), esp. RACK-1 (R85850). Proteins R85851-82 were isolated based on
CC homology with beta-transducin, whereas proteins R85882-92 were isolated
CC based on homology with the WD-40 consensus sequence (R85893). The

```
CC   proteins were used to construct the peptides R84928-R85063 and
CC   R85786-R85842.  The peptides can be used to identify target proteins
CC   contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC   proteins involved in protein-protein interaction and to screen for drugs
CC   that will affect protein-protein interaction involving WD-40 domains.
SQ   Sequence   603 AA;

  Query Match          24.9%;  Score 242;  DB 17;  Length 603;
  Best Local Similarity  34.1%;  Pred. No. 7.10e-09;
  Matches   46;  Conservative   34;  Mismatches  52;  Indels   3;  Gaps   3;

Db    240 hlprlqklyldrnlitavapgaflgmkalrwldlshnrvaglmedtfpgllglhvlrlah 299
          ||  || | :: | |::: |||  :| | :| |::| :  : |  | |   |  | |:
Qy      1 HLRVLQ-L-ME-NRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSE 57

Db    300 naiaslrprtfkdlhfleelqlghnrirqlgertfeglgqlevltlndnqitevrvgafs 359
          | | ::  ::|:    : :|||  |:|  : : :| :| :|||||||:| || : |::|:
Qy     58 NQIQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFN 117

Db    360 glfnvavmnlsgncl 374
           : ::    | :| |
Qy    118 HMPKLRTFRLHSNNL 132


RESULT    7
ID   R05160 standard; protein; 353 AA.
AC   R05160;
DT   09-OCT-1990  (first entry)
DE   Sequence of human bone proteoglycan II (decorin).
KW   Osteoporosis; rheumatoid arthritis; Paget's disease;
KW   atherosclerosis; periodontal; human bone matrix; proteoglycan.
OS   Homo sapiens.
PN   US7432044-A.
PD   17-APR-1990.
PF   3-NOV-1989; 432044.
PR   3-NOV-1989; US-432044.
PA   (USSH) Nat Inst of Health.
PI   Termine J;
DR   WPI; 90-178641/23.
DR   N-PSDB; Q04491.
PT   Human bone matrix DNA and proteins -
PT   used in detection, diagnosis and treatment involving skeletal
PT   and/or connective tissue disease states.
PS   Disclosure; p; English.
CC   Probes and Abs raised to the proteins can be used to determine
CC   their levels useful in diagnosis of associated conective tissue
CC   diseases states such as osteoporosis, osteo/rheumatoid arthritis,
CC   Paget's disease, artherosclerosis and periodontal disease.
CC   Proteins may also be used to induce or block biological function.
SQ   Sequence   353 AA;

  Query Match          22.7%;  Score 220;  DB 1;  Length 353;
  Best Local Similarity  32.8%;  Pred. No. 2.56e-07;
  Matches   41;  Conservative   33;  Mismatches  46;  Indels   5;  Gaps   4;

Db    100 nlhalilvnnkiskvspgaftplvklerlylsknqlkelpekm--pkt-lqelrahenei 156
          :|: | |::|:|| :  |||  |  |||| |::| |  :|| :    : |  |   ||:|
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    157 tkvrkvtfnglnqmivielgtnplkssgiengafqgmkklsyiriadtnitkvrkvtfng 216
            : : :| |  ::  ::|  | : |  ||:|||:::: |  : : : |||::   :||
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQI-SC-IEDGAFRALRDLEVLTLNNNNITRLSVASFNH 118

Db    217 ltelh 221
          :  |:
Qy    119 MPKLR 123


RESULT    8
ID   R87952 standard; Protein; 369 AA.
AC   R87952;
DT   20-MAR-1996  (first entry)
DE   Human neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..37
FT                   /label= Sig_peptide
FT   region          44..60
FT                   /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s) -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 3; Fig 8; 60pp; English.
CC   Human biglycan (R87952) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  It can be used to enhance
CC   the survival and maintain the structure and function of CNS neurons
CC   during normal ageing as well as after pathological and/or traumatic
CC   nervous system damage.  It can also be used to restore function
CC   following nervous system lesions and degenerative diseases, and to
CC   improve learning efficiency and memory in the elderly and in patients
CC   with dementia.
SQ   Sequence   369 AA;

  Query Match          22.1%;  Score 215;  DB 15;  Length 369;
  Best Local Similarity  36.4%;  Pred. No. 5.75e-07;
  Matches   43;  Conservative   29;  Mismatches  41;  Indels   5;  Gaps   5;

Db    209 lklnylriseakltgipkdlpetlnelhldhnkiqai-eledllrysklyrlglghnqir 267
          | |   |||     |   || | |: |:|::|::| : ||  :|  ::|||| |: |||:
Qy      5 LQLMENRISTIE-RGAFQDLKE-LERLRLNRNNLQLFPELL-FLGTARLYRLDLSENQIQ 61

Db    268 miengslsflptlrelhldnnklsrvpag-lpdlkllqvvylhsnnitkvgvndfcpm 324
            |   ::     :::|:|| | :| :  | :  |: |:|: |::||||:::|  |  |
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119


RESULT    9
ID   R05159 standard; protein; 368 AA.
AC   R05159;
DT   09-OCT-1990  (first entry)
DE   Sequence of human bone proteoglycan I (biglycan).
KW   Osteoporosis; rheumatoid arthritis; Paget's disease;
KW   atherosclerosis; periodontal; human bone matrix; proteoglycan.
OS   Homo sapiens.
PN   US7432044-A.
PD   17-APR-1990.
PF   3-NOV-1989; 432044.
PR   3-NOV-1989; US-432044.
PA   (USSH) Nat Inst of Health.
PI   Termine J;
DR   WPI; 90-178641/23.
DR   N-PSDB; Q04490.
PT   Human bone matrix DNA and proteins -
PT   used in detection, diagnosis and treatment involving skeletal
PT   and/or connective tissue disease states.
PS   Disclosure; p; English.
CC   Probes and Abs raised to the proteins can be used to determine
CC   their levels useful in diagnosis of associated conective tissue
CC   diseases states such as osteoporosis, osteo/rheumatoid arthritis,
CC   Paget's disease, artherosclerosis and periodontal disease.
CC   Proteins may also be used to induce or block biological function.
SQ   Sequence   368 AA;

  Query Match          21.9%;  Score 213;  DB 1;  Length 368;
  Best Local Similarity  35.6%;  Pred. No. 7.94e-07;
```

```
  Matches   42;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   5;

Db    208 lklnylriseakltgipkdlpetlnelhldhnkiqai-eledllrysklyrlglghnqir 266
          | |   |||     |   || | |: |:|::|::| : ||  :|  ::|||| |: |||:
Qy      5 LQLMENRISTIE-RGAFQDLKE-LERLRLNRNNLQLFPELL-FLGTARLYRLDLSENQIQ 61

Db    267 miengslsflptlrelhldnnklarvpsg-lpdlkllqvvylhsnnitkvgvndfcpm 323
          |   ::     :::|:|| | :: :  | :  |: |:|: |::||||:::|  | |
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119


RESULT   10
ID   R87951 standard; Protein; 369 AA.
AC   R87951;
DT   20-MAR-1996  (first entry)
DE   Rat neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   peptide         1..37
FT                   /label= Sig_peptide
FT   region          44..60
FT                   /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
DR   N-PSDB; T08768.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)  -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 1; Page 44-45; 60pp; English.
CC   Rat biglycan (R87951) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  Recombinant biglycan,
CC   obtd. by expression of encoding cDNA (T08768) in eukaryotic host
CC   cells, can be used to enhance the survival and maintain the structure
CC   and function of CNS neurons during normal ageing as well as after
CC   pathological and/or traumatic nervous system damage.  It can also
CC   be used to restore function following nervous system lesions and
CC   degenerative diseases, and to improve learning efficiency and memory
CC   in the elderly and in patients with dementia.
SQ   Sequence   369 AA;

  Query Match           21.9%;  Score 213;  DB 15;  Length 369;
  Best Local Similarity  35.6%;  Pred. No. 7.94e-07;
  Matches   42;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   5;

Db    209 lklnylriseakltgipkdlpetlnelhldhnkiqai-eledllrysklyrlglghnqir 267
          | |   |||     |   || | |: |:|::|::| : ||  :|  ::|||| |: |||:
Qy      5 LQLMENRISTIE-RGAFQDLKE-LERLRLNRNNLQLFPELL-FLGTARLYRLDLSENQIQ 61

Db    268 miengslsflptlrelhldnnklsrvpag-lpdlkllqvvylhsnnitkvgindfcpm 324
          |   ::     :::|:|| | :|  :  | :  |: |:|: |::||||:::  | |
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119


RESULT   11
ID   R87953 standard; Protein; 332 AA.
AC   R87953;
DT   20-MAR-1996  (first entry)
DE   Bovine neurotrophic biglycan.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   region          7..23
FT                   /label= Hypervariable_region
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)  -
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Claim 3; Fig 8; 60pp; English.
CC   Bovine biglycan (R87953) is a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  It can be used to enhance
CC   the survival and maintain the structure and function of CNS neurons
CC   during normal ageing as well as after pathological and/or traumatic
CC   nervous system damage.  It can also be used to restore function
CC   following nervous system lesions and degenerative diseases, and to
CC   improve learning efficiency and memory in the elderly and in patients
CC   with dementia.
SQ   Sequence   332 AA;

  Query Match           21.6%;  Score 210;  DB 15;  Length 332;
  Best Local Similarity  36.5%;  Pred. No. 1.29e-06;
  Matches   42;  Conservative   28;  Mismatches  40;  Indels   5;  Gaps   5;

Db    172 lklnylriseakltgipkdlpetlnelhldhnkiqai-eledllrysklyrlglghnqir 230
          | |   |||     |   || | |: |:|::|::| : ||  :|  ::|||| |: |||:
Qy      5 LQLMENRISTIE-RGAFQDLKE-LERLRLNRNNLQLFPELL-FLGTARLYRLDLSENQIQ 61

Db    231 miengslsflptlrelhldnnklsrvpag-lpdlkllqvvylhtnnitkvgvndf 284
          |   ::     :::|:|| | :| :  | :  |: |:|: |: ||||:::|  |
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASF 116


RESULT   12
ID   R42265 standard; Protein; 234 AA.
AC   R42265;
DT   28-APR-1994  (first entry)
DE   Decorin sequence PT-76 (N-terminal to LRR8).
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   N-PSDB; Q50051.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Claim 10; Page 45-46; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   234 AA;

  Query Match           20.2%;  Score 196;  DB 8;  Length 234;
  Best Local Similarity  33.3%;  Pred. No. 1.22e-05;
  Matches   38;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   4;

Db     78 nlhalilvnnkiskvspgaftplvklerlylsknqlkelpekm--pkt-lqelrahenei 134
          :|: | |::|:|| :  |||  |  |||| |::| |  :|| :    : |  |   ||:|
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60
```

```
Db    135 tkvrkvtfnglnqmivielgtnplkssgiengafqgmkklsyiriadtnitsip 188
          : : :| |  ::  ::|  | : |  ||:|||:::: |  : : : ||| ::
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQI-SC-IEDGAFRALRDLEVLTLNNNNITRLS 112


RESULT  13
ID   R42266 standard; Protein; 280 AA.
AC   R42266;
DT   28-APR-1994  (first entry)
DE   Decorin sequence PT-77 (N-terminal to LRR10).
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   N-PSDB; Q50052.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Claim 10; Page 47-48; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   280 AA;

  Query Match           20.2%;  Score 196;  DB 8;  Length 280;
  Best Local Similarity  33.3%;  Pred. No. 1.22e-05;
  Matches   38;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   4;

Db     78 nlhalilvnnkiskvspgaftplvklerlylsknqlkelpekm--pkt-lqelrahenei 134
          :|: | |::|:|| :  |||  |  |||| |::| |  :|| :    : |  |   ||:|
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    135 tkvrkvtfnglnqmivielgtnplkssgiengafqgmkklsyiriadtnitsip 188
          : : :| |  ::  ::|  | : |  ||:|||:::: |  : : : ||| ::
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQI-SC-IEDGAFRALRDLEVLTLNNNNITRLS 112


RESULT  14
ID   R42267 standard; Protein; 305 AA.
AC   R42267;
DT   28-APR-1994  (first entry)
DE   Decorin sequence PT-78 (N-terminal to half C-terminal).
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   N-PSDB; Q50053.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Claim 10; Page 49-50; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   305 AA;

  Query Match           20.2%;  Score 196;  DB 8;  Length 305;
  Best Local Similarity  33.3%;  Pred. No. 1.22e-05;
  Matches   38;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   4;

Db     78 nlhalilvnnkiskvspgaftplvklerlylsknqlkelpekm--pkt-lqelrahenei 134
          :|: | |::|:|| :  |||  |  |||| |::| |  :|| :    : |  |   ||:|
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    135 tkvrkvtfnglnqmivielgtnplkssgiengafqgmkklsyiriadtnitsip 188
          : : :| |  ::  ::|  | : |  ||:|||:::: |  : : : ||| ::
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQI-SC-IEDGAFRALRDLEVLTLNNNNITRLS 112

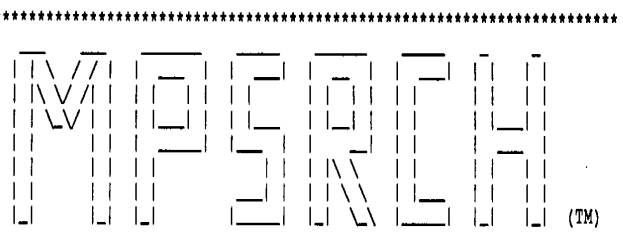

RESULT  15
ID   R42260 standard; Protein; 331 AA.
AC   R42260;
DT   28-APR-1994  (first entry)
DE   Mature decorin PT-65.
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta.
FH   Key             Location/Qualifiers
FT   region          1..45
FT                   /label= N-terminal_region
FT                   /note= "contains 4 Cys residues"
FT   region          46..280
FT                   /label= repeat_region
FT                   /note= "contains 10 leucine-rich repeats"
FT   region          281..331
FT                   /label= C-terminal_region
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   N-PSDB; Q50046.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Claim 10; Page 36-38; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   331 AA;

  Query Match           20.2%;  Score 196;  DB 8;  Length 331;
  Best Local Similarity  33.3%;  Pred. No. 1.22e-05;
  Matches   38;  Conservative   30;  Mismatches  41;  Indels   5;  Gaps   4;

Db     78 nlhalilvnnkiskvspgaftplvklerlylsknqlkelpekm--pkt-lqelrahenei 134
          :|: | |::|:|| :  |||  |  |||| |::| |  :|| :    : |  |   ||:|
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    135 tkvrkvtfnglnqmivielgtnplkssgiengafqgmkklsyiriadtnitsip 188
          : : :| |  ::  ::|  | : |  ||:|||:::: |  : : : ||| ::
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQI-SC-IEDGAFRALRDLEVLTLNNNNITRLS 112


Search completed: Fri May 28 09:29:40 1999
Job time : 58 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:29:57 1999; MasPar time 8.33 Seconds
644.624 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-12
Description: (1-134) from US09191647.pep
Perfect Score: 971
Sequence: 1 HLRVLQLMENRISTIERGAF.....SFNHMPKLRTFRLHSNNLYC 134

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 44.958; Variance 121.862; scale 0.369

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	366	37.7	1469	2	B36665	slit protein 2 precur	7.70e-34
2	366	37.7	1480	2	A36665	slit protein 1 precur	7.70e-34
3	298	30.7	682	2	A49121	cell-surface molecule	2.53e-24
4	298	30.7	682	2	A43318	connectin precursor -	2.53e-24
5	267	27.5	1091	2	A58532	glial cell membrane g	4.31e-20
6	264	27.2	1134	1	A29944	chaoptin precursor -	1.10e-19
7	261	26.9	907	2	JE0176	orphan G protein-coup	2.78e-19
8	260	26.8	536	2	A34901	lysine carboxypeptida	3.79e-19
9	258	26.6	560	2	A60164	platelet membrane gly	7.04e-19
10	252	26.0	605	2	A41915	insulin-like growth f	4.48e-18
11	248	25.5	605	2	JC5239	insulin-like growth f	1.53e-17
12	242	24.9	603	2	JC1282	insulin-like growth f	9.62e-17
13	241	24.8	603	2	JC6128	insulin-like growth f	1.31e-16
14	237	24.4	1115	2	S40241	G protein-coupled rec	4.41e-16
15	235	24.2	312	1	NBHUA2	leucine-rich alpha-2-	8.10e-16
16	222	22.9	361	2	A53860	chondroadherin precur	4.09e-14
17	222	22.9	662	2	S42799	garp precursor - huma	4.09e-14
18	213	21.9	368	1	BGHUN	biglycan precursor -	6.00e-13
19	213	21.9	369	2	S32793	biglycan precursor -	6.00e-13
20	213	21.9	369	2	S20811	proteoglycan I - mous	6.00e-13
21	210	21.6	369	2	S32559	biglycan precursor -	1.46e-12
22	208	21.4	1097	2	A29943	Toll protein precurso	2.63e-12
23	205	21.1	357	2	S24317	decorin precursor - c	6.38e-12
24	203	20.9	1535	2	S46224	peroxidasin - fruit f	1.15e-11
25	202	20.8	360	2	S06280	decorin precursor - b	1.54e-11
26	201	20.7	360	2	I47020	decorin - rabbit	2.06e-11
27	196	20.2	359	1	NBHUC8	decorin precursor - h	8.87e-11
28	192	19.8	1256	2	S60461	gene flightless-I pro	2.83e-10
29	190	19.6	354	2	A55454	decorin precursor - m	5.04e-10
30	189	19.5	382	2	I39068	proline- arginine-ric	6.72e-10
31	188	19.4	925	2	JC2033	G protein-coupled rec	8.96e-10
32	182	18.7	424	2	S27783	hypothetical protein	4.99e-09
33	180	18.5	354	2	S29145	decorin precursor - r	8.81e-09
34	180	18.5	626	1	NBHUIA	platelet glycoprotein	8.81e-09
35	180	18.5	661	2	I56258	RP105 - mouse	8.81e-09
36	179	18.4	277	2	S25770	RSP-1 protein - mouse	1.17e-08
37	178	18.3	277	2	I60122	rsu-1 homolog - human	1.55e-08
38	176	18.1	1268	2	A49674	flightless-I homolog	2.73e-08
39	175	18.0	1839	1	OYBYK	adenylate cyclase (EC	3.62e-08
40	172	17.7	376	2	S55275	fibromodulin precurso	8.42e-08
41	171	17.6	2145	2	JC4747	adenylate cyclase (EC	1.11e-07
42	165	17.0	375	2	S05390	fibromodulin precurso	5.90e-07
43	162	16.7	764	2	JC5643	thyroid stimulating h	1.35e-06
44	159	16.4	2026	1	OYBY	adenylate cyclase (EC	3.07e-06
45	157	16.2	1692	2	A33988	adenylate cyclase (EC	5.29e-06

ALIGNMENTS

RESULT 1
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 16-Dec-1998
ACCESSIONS B36665
REFERENCE A36665
#authors Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
#journal Genes Dev. (1990) 4:2169-2187
#title slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.
#cross-references MUID:91099665
#accession B36665
##status preliminary
##molecule_type mRNA
##residues 1-1469 ##label ROT
##cross-references GB:X53959
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
CLASSIFICATION #superfamily proteoglycan amino-terminal homology; EGF homology; leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-terminal homology
FEATURE
66-91 #domain proteoglycan amino-terminal homology #label PAH1\
101-124 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR1\
125-148 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR2\
149-172 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR3\
173-196 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR4\
197-220 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
228-272 #domain proteoglycan carboxyl-terminal homology #label PCS1\
288-313 #domain proteoglycan amino-terminal homology #label PAH2\
323-346 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
347-370 #domain leucine-rich alpha-2-glycoprotein repeat

```
                  homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           37.7%;  Score 366;  DB 2;  Length 1469;
  Best Local Similarity  40.0%;  Pred. No. 7.70e-34;
  Matches   52;  Conservative   31;  Mismatches  47;  Indels   0;  Gaps   0;

Db    105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
          |:|  | :: |    || |  |  |:|  | :: :    |     | |||:|:| | ::
Qy      5 LQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQAIP 64

Db    165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
          |: |:||  :::|||| |||:|::: ||::| :||:|||||||:| |:   |  : :||:
Qy     65 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 124

Db    225 LRLSDNPFAC 234
          :||  | : |
Qy    125 FRLHSNNLYC 134


ULT    2
RY            A36665     #type complete
ITLE          slit protein 1 precursor - fruit fly (Drosophila
                melanogaster)
ORGANISM      #formal_name Drosophila melanogaster
DATE          30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                24-Sep-1998
ACCESSIONS    A36665; S13523
REFERENCE     A36665
  #authors    Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                Artavanis-Tsakonas, S.
  #journal    Genes Dev. (1990) 4:2169-2187
  #title      slit: an extracellular protein necessary for development of
                midline glia and commissural axon pathways contains both
                EGF and LRR domains.
  #cross-references MUID:91099665
  #accession  A36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1480 ##label ROT
    ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
  #gene       FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                homology; leucine-rich alpha-2-glycoprotein repeat
                homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
   66-91          #domain proteoglycan amino-terminal homology #label
                  PAH1\
   101-124        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR1\
   125-148        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR2\
   149-172        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR3\
   173-196        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR4\
   197-220        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR5\
   228-272        #domain proteoglycan carboxyl-terminal homology #label
                  PCS1\
   288-313        #domain proteoglycan amino-terminal homology #label
                  PAH2\
   323-346        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR6\
   347-370        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR7\
   371-394        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR8\
   395-418        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LRR9\
   419-442        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR10\
   450-494        #domain proteoglycan carboxyl-terminal homology #label
                  PCS2\
   512-537        #domain proteoglycan amino-terminal homology #label
                  PAH3\
   547-571        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR11\
   572-595        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR12\
   596-619        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR13\
   620-643        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
   651-695        #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
   708-733        #domain proteoglycan amino-terminal homology #label
                  PAH4\
   743-766        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
   767-790        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
   791-814        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
   815-838        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
   846-890        #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
   1028-1061      #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           37.7%;  Score 366;  DB 2;  Length 1480;
  Best Local Similarity  40.0%;  Pred. No. 7.70e-34;
  Matches   52;  Conservative   31;  Mismatches  47;  Indels   0;  Gaps   0;

Db    105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
          |:|  | :: |    || |  |  |:|  | :: :    |     | |||:|:| | ::
Qy      5 LQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQAIP 64

Db    165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
          |: |:||  :::|||| |||:|::: ||::| :||:|||||||:| |:   |  : :||:
Qy     65 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 124

Db    225 LRLSDNPFAC 234
```

```
        :||  | : |
Qy    125 FRLHSNNLYC 134


RESULT    3
ENTRY           A49121     #type complete
TITLE           cell-surface molecule connectin - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                  20-Mar-1998
ACCESSIONS      A49121
REFERENCE       A49121
   #authors     Gould, A.P.; White, R.A.
   #journal     Development (1992) 116:1163-1174
   #title       Connectin, a target of homeotic gene control in Drosophila.
   #cross-references MUID:93202002
   #accession   A49121
      ##status       preliminary
      ##molecule_type nucleic acid
      ##residues     1-682 ##label GOU
      ##cross-references GB:X68701; NID:g7737; PID:g7738
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIN:127661,
                      NCBIP:127664)
GENETICS
   #gene        FlyBase:Con
      ##cross-references FlyBase:FBgn0005775
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
FEATURE
   199-222          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   223-246          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   247-270          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   271-294          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   295-318          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   319-342          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   343-366          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7
SUMMARY         #length 682  #molecular-weight 75922  #checksum 7093

  Query Match           30.7%;  Score 298;  DB 2;  Length 682;
  Best Local Similarity  38.3%;  Pred. No. 2.53e-24;
  Matches   51;  Conservative   28;  Mismatches  52;  Indels   2;  Gaps   2;

Db    223 RLRELNLEHNQIFEMDRYAFRNL-PLCERLFLNNNNISTLHEGLFADMARLTFLNLAHNQ 281
          :|| | |  |:|  ::| ||::|  | ||| || ||:  : | ||   |||  |:|: ||
Qy      1 HLRVLQLMENRISTIERGAFQDLKEL-ERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db    282 INVLTSEIFRGLGNLNVLKLTRNNLNFIGDTVFAELWSLSELELDDNRIERISERALDGL 341
          |  :   |||  ::: | |  | :: | |  |  |: |  | |::| | |:|  ::: :
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    342 NTLKTLNLRNNLL 354
           |:|: |::| |
Qy    120 PKLRTFRLHSNNL 132


RESULT    4
ENTRY           A43318     #type complete
TITLE           connectin precursor - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
                  24-Sep-1998
ACCESSIONS      A43318; S28464
REFERENCE       A43318
   #authors     Nose, A.; Mahajan, V.B.; Goodman, C.S.
   #journal     Cell (1992) 70:553-567
   #title       Connectin: a homophilic cell adhesion molecule expressed on a
                  subset of muscles and the motoneurons that innervate them
                  in Drosophila.
   #cross-references MUID:92370678
   #accession   A43318
      ##molecule_type mRNA
      ##residues     1-682 ##label NOS
      ##cross-references GB:M96647; NID:g157083; PID:g157084
      ##note         sequence extracted from NCBI backbone (NCBIN:111422,
                      NCBIP:111423)
REFERENCE       S28464
   #authors     Gould, A.P.; White, R.A.H.
   #submission  submitted to the EMBL Data Library, October 1992
   #description Connectin a target of homeotic gene control in drosophila.
   #accession   S28464
      ##molecule_type mRNA
      ##residues     1-630,'G',632-673,675-678,'M',679-682 ##label GOU
      ##cross-references EMBL:X68701; NID:g7737; PID:g7738
GENETICS
   #gene        FlyBase:Con
      ##cross-references FlyBase:FBgn0005775
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
FEATURE
   1-26             #domain signal sequence #status predicted #label SIG\
   27-682           #product connectin #status predicted #label MAT
SUMMARY         #length 682  #molecular-weight 75991  #checksum 7269

  Query Match           30.7%;  Score 298;  DB 2;  Length 682;
  Best Local Similarity  38.3%;  Pred. No. 2.53e-24;
  Matches   51;  Conservative   28;  Mismatches  52;  Indels   2;  Gaps   2;

Db    223 RLRELNLEHNQIFEMDRYAFRNL-PLCERLFLNNNNISTLHEGLFADMARLTFLNLAHNQ 281
          :|| | |  |:|  ::| ||::|  | ||| || ||:  : | ||   |||  |:|: ||
Qy      1 HLRVLQLMENRISTIERGAFQDLKEL-ERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db    282 INVLTSEIFRGLGNLNVLKLTRNNLNFIGDTVFAELWSLSELELDDNRIERISERALDGL 341
          |  :   |||  ::: | |  | :: | |  |  |: |  | |::| | |:|  ::: :
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    342 NTLKTLNLRNNLL 354
           |:|: |::| |
Qy    120 PKLRTFRLHSNNL 132


RESULT    5
ENTRY           A58532     #type complete
TITLE           glial cell membrane glycoprotein LIG-1 precursor - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change
                  17-Mar-1999
ACCESSIONS      A58532
REFERENCE       A58532
   #authors     Suzuki, Y.; Sato, N.; Tohyama, M.; Wanaka, A.; Takagi, T.
   #journal     J. Biol. Chem. (1996) 271:22522-22527
   #title       cDNA cloning of a novel membrane glycoprotein that is
                  expressed specifically in glial cells in the mouse brain;
                  LIG-1, a protein with leucine-rich repeats and
                  immunoglobulin-like domains.
   #cross-references MUID:96394313
   #accession   A58532
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-1091 ##label SUZ
      ##cross-references GB:D78572; NID:g1545806; PID:g1545807
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan amino-terminal homology;
                  proteoglycan carboxyl-terminal homology
FEATURE
   36-61            #domain proteoglycan amino-terminal homology #label PAH\
   71-94            #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
```

```
   95-117          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR2\
   118-141         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR3\
   142-165         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR4\
   166-189         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR5\
   191-213         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR6\
   214-237         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR7\
   238-261         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR8\
   262-285         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR9\
   286-309         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR10\
   310-333         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR11\
   334-357         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR12\
   358-381         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR13\
   385-408         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR14\
   409-432         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR15\
   440-485         #domain proteoglycan carboxyl-terminal homology #label
                     PCS1
SUMMARY          #length 1091  #molecular-weight 119283  #checksum 7937

  Query Match           27.5%;  Score 267;  DB 2;  Length 1091;
  Best Local Similarity  35.3%;  Pred. No. 4.31e-20;
  Matches   47;  Conservative   37;  Mismatches  47;  Indels   2;  Gaps   2;

Db    166 RIRELNLASNRISILESGAFDGLSRSLLTLRLSKNRITQLPVKAF-KLPRLTQLDLNRNR 224
          ::| | |  |||| :| |||: | : |  |||::| :  :|   |   :|| :|||: |:
Qy      1 HLRVLQLMENRISTIERGAFQDL-KELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db    225 IRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVNSGSLYGL 284
          |: |   :|:|   :  |:|: | || : ||||::|  : || |: |::  :: :|:  :
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    285 TALHQLHLSNNSI 297
            |: ::| :|::
Qy    120 PKLRTFRLHSNNL 132


ULT    6
RY             A29944     #type complete
TITLE          chaoptin precursor - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  photoreceptor cell-specific membrane protein
ORGANISM       #formal_name Drosophila melanogaster
DATE           15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change
                 24-Oct-1997
ACCESSIONS     A29944; A21123
REFERENCE      A29944
   #authors    Reinke, R.; Krantz, D.E.; Yen, D.; Zipursky, S.L.
   #journal    Cell (1988) 52:291-301
   #title      Chaoptin, a cell surface glycoprotein required for Drosophila
                photoreceptor cell morphogenesis, contains a repeat motif
                found in yeast and human.
   #cross-references MUID:88135762
   #accession  A29944
      ##molecule_type DNA
      ##residues     1-1134 ##label REI
      ##cross-references GB:M19008; GB:M19009; GB:M19010; GB:M19011;
                        GB:M19012; GB:M19013; GB:M19014; GB:M19015;
                        GB:M19016; GB:M19017; NID:g157094; PID:g157098
REFERENCE      A21123
   #authors    Zipursky, S.L.; Venkatesh, T.R.; Teplow, D.B.; Benzer, S.
   #journal    Cell (1984) 36:15-26
   #title      Neuronal development in the Drosophila retina: monoclonal
                antibodies as molecular probes.
   #cross-references MUID:84106810
   #accession  A21123
      ##molecule_type protein
      ##residues     31-43,'HX',46-49,'H' ##label ZIP
GENETICS
   #gene       FlyBase:chp
      ##cross-references FlyBase:FBgn0000313
   #introns    1/3 80/3; 318/3; 377/2 422/2; 702/1 745/3; 831/2; 998/2
CLASSIFICATION  #superfamily chaoptin; leucine-rich alpha-2-glycoprotein
                 repeat homology
KEYWORDS       cell adhesion; glycoprotein; membrane protein
FEATURE
   1-29            #domain signal sequence #status predicted #label SIG\
   30-1134         #product chaoptin #status predicted #label MAT\
   80-102          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #status atypical #label LRR1\
   103-126         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR2\
   128-151         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR3\
   152-175         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR4\
   177-200         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR5\
   201-224         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR6\
   226-249         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR7\
   250-273         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR8\
   279-302         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LRR9\
   303-325         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR10\
   326-349         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR11\
   351-374         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR12\
   375-399         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR13\
   401-424         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR14\
   428-451         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR15\
   453-476         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR16\
   477-500         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR17\
   502-526         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR18\
   527-550         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR19\
   551-574         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR20\
   577-600         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR21\
   601-624         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR22\
   625-648         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR23\
   649-672         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR24\
   673-696         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR25\
   708-731         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR26\
   733-756         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR27\
   757-780         #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR28\
   781-804         #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LR29\
   805-827        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR30\
   828-851        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR31\
   854-877        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR32\
   879-902        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR33\
   903-926        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR34\
   928-948        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #status atypical #label LR35\
   949-972        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR36\
   973-995        #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR37\
   996-1019       #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR38\
   1021-1044      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR39\
   1056-1080      #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR40
SUMMARY         #length 1134  #molecular-weight 130719  #checksum 332

  Query Match           27.2%;  Score 264;  DB 1;  Length 1134;
  Best Local Similarity  35.6%;  Pred. No. 1.10e-19;
  Matches   48;  Conservative   35;  Mismatches  47;  Indels   5;  Gaps   5;

Db     80 KVFMLHMENTGLREIEP-YFLQSTGMYRLKISGNHLTEIPDDAFTGLERSLWELILPQND 138
          :|: | |||  :  ||   | :   : ||:::|:|  :|:  | |  | |  | |::|:
Qy      3 RVLQL-MENR-ISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTAR-LYRLDLSENQ 59

Db    139 LVEIPSKSLRHLQKLRHLDLGYNHITHIQHDSFRGLEDSLQTLILRENCISQLMSHSFSG 198
          :  || |::|    :::|:| ||:|: |:  :||:| | |: | | :| |::|   ||:
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRD-LEVLTLNNNNITRLSVASFNH 118

Db    199 LLILETLDLSGNNLF 213
          :  | |: | :|||:
Qy    119 MPKLRTFRLHSNNLY 133


RESULT    7
ENTRY           JE0176     #type complete
TITLE           orphan G protein-coupled receptor precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      JE0176
REFERENCE       JE0176
   #authors     McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.;
                  Caskey, C.T.; Liu, Q.
   #journal     Biochem. Biophys. Res. Commun. (1998) 247:266-270
   #title       Identification and cloning of an orphan G protein-coupled
                  receptor of the glycoprotein hormone receptor subfamily.
   #cross-references MUID:98308104
   #accession   JE0176
      ##molecule_type mRNA
      ##residues      1-907 ##label MCD
      ##cross-references GB:AF062006
COMMENT    This protein is a receptor for a novel class of glycoprotein
             ligands.
GENETICS
   #gene        HG38
   #map_position 12q22-23
FEATURE
   1-21             #domain signal sequence #status predicted #label SIG\
   562-583          #domain transmembrane #status predicted #label TM1\
   594-616          #domain transmembrane #status predicted #label TM2\
   639-660          #domain transmembrane #status predicted #label TM3\
   681-701          #domain transmembrane #status predicted #label TM4\
   725-744          #domain transmembrane #status predicted #label TM5\
   768-791          #domain transmembrane #status predicted #label TM6\
   803-824          #domain transmembrane #status predicted #label TM7
SUMMARY         #length 907  #molecular-weight 99997  #checksum 8790

  Query Match           26.9%;  Score 261;  DB 2;  Length 907;
  Best Local Similarity  36.6%;  Pred. No. 2.78e-19;
  Matches   48;  Conservative   24;  Mismatches  59;  Indels   0;  Gaps   0;

Db    116 LKVLMLQNNQLRHVPTEALQNLRSLQSLRLDANHISYVPPSCFSGLHSLRHLWLDDNALT 175
          |:|| | :|::  :   |:|:|: |: |||: |::   |   | |   | :| | :| :
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    176 EIPVQAFRSLSALQAMTLALNKIHHIPDYAFGNLSSLVVLHLHNNRIHSLGKKCFDGLHS 235
            ||  |||:   :  : |  | |  | | ||  |  | | || |:|| |  |:   |: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    236 LETLDLNYNNL 246
          | |: |: |||
Qy    122 LRTFRLHSNNL 132


RESULT    8
ENTRY           A34901     #type complete
TITLE           lysine carboxypeptidase (EC 3.4.17.3) 83K chain - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change
                  24-Sep-1998
ACCESSIONS      A34901
REFERENCE       A34901
   #authors     Tan, F.; Weerasinghe, D.K.; Skidgel, R.A.; Tamei, H.; Kaul,
                  R.K.; Roninson, I.B.; Schilling, J.W.; Erdoes, E.G.
   #journal     J. Biol. Chem. (1990) 265:13-19
   #title       The deduced protein sequence of the human carboxypeptidase N
                  high molecular weight subunit reveals the presence of
                  leucine-rich tandem repeats.
   #cross-references MUID:90094386
   #accession   A34901
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-536 ##label TAN
      ##cross-references GB:J05158; NID:g179935; PID:g179936
GENETICS
   #gene        GDB:ACBP
      ##cross-references GDB:127893
   #map_position 6q25.3-6q26
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        hydrolase; metallo-carboxypeptidase
FEATURE
   77-100           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   101-124          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   125-148          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   149-172          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   173-196          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   197-220          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   221-244          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   245-268          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   269-292          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   293-316          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   317-340          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   341-364          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12
SUMMARY         #length 536  #molecular-weight 58649  #checksum 8569
```

```
  Query Match           26.8%;  Score 260;  DB 2;  Length 536;
  Best Local Similarity  32.6%;  Pred. No. 3.79e-19;
  Matches   43;  Conservative   31;  Mismatches  58;  Indels   0;  Gaps   0;

Db     77 RLEDLEVTGSSFLNLSTNIFSNLTSLGKLTLNFNMLEALPEGLFQHLAALESLHLQGNQL 136
          :|  |::  : :  :    | :|  | :| || | |: :|| ||   | |  | |  ||:
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    137 QALPRRLFQPLTHLKTLNLAQNLLAQLPEELFHPLTSLQTLKLSNNALSGLPQGVFGKLG 196
          ||:||: |:    :| | |  | :: : :  |::|  |: | |:|| :: |: : |  :
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    197 SLQELFLDSNNI 208
           |: : | |||:
Qy    121 KLRTFRLHSNNL 132


RESULT    9
ENTRY           A60164     #type complete
TITLE           platelet membrane glycoprotein V precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            12-Jan-1993 #sequence_revision 24-Feb-1994 #text_change
                  17-Mar-1999
ACCESSIONS      A48030; A60164; A35483; B35483; C35483; A60432; A47507;
                  S34329
REFERENCE       A48030
  #authors      Lanza, F.; Morales, M.; de La Salle, C.; Cazenave, J.P.;
                  Clemetson, K.J.; Shimomura, T.; Phillips, D.R.
  #journal      J. Biol. Chem. (1993) 268:20801-20807
  #title        Cloning and characterization of the gene encoding the human
                  platelet glycoprotein V. A member of the leucine-rich
                  glycoprotein family cleaved during thrombin-induced
                  platelet activation.
  #cross-references MUID:94012616
  #accession    A48030
    ##molecule_type DNA
    ##residues      1-560 ##label LA2
    ##cross-references EMBL:Z23091; NID:g312501; PID:g312502
REFERENCE       A60164
  #authors      Shimomura, T.; Fujimura, K.; Maehama, S.; Takemoto, M.; Oda,
                  K.; Fujimoto, T.; Oyama, R.; Suzuki, M.; Ichihara-Tanaka,
                  K.; Titani, K.; Kuramoto, A.
  #journal      Blood (1990) 75:2349-2356
  #title        Rapid purification and characterization of human platelet
                  glycoprotein V: the amino acid sequence contains
                  leucine-rich repetitive modules as in glycoprotein Ib.
  #cross-references MUID:90275263
  #accession    A60164
    ##molecule_type protein
    ##residues      365-384,'X',386-390,'X',392-395,'X',397;188-208,'I',210;
                     27-50,'X',52-53;174-180,'X',182-187;121-144;145-172;
                     290-297,'X',299-311,'X',313-326,'I';142-151,'X',
                     153-163;'YNTPDRXLAXYGGF';81-105,'XX',108,'T';61-72,
                     'TK',75-77;'V',56-57;'G',479-487,'X',489-498,'X',500,
                     'X',502-503,'X',505,'X',507-508,'D' ##label SHI
REFERENCE       A35483
  #authors      Roth, G.J.; Church, T.A.; McMullen, B.A.; Williams, S.A.
  #journal      Biochem. Biophys. Res. Commun. (1990) 170:153-161
  #title        Human platelet glycoprotein V: a surface leucine-rich
                  glycoprotein related to adhesion.
  #cross-references MUID:90321220
  #accession    A35483
    ##molecule_type protein
    ##residues      145-166,'I',168-169,'X',171-172 ##label ROT
    ##note          this proteolytic fragment was designated peptide M392
  #accession    B35483
    ##molecule_type protein
    ##residues      121-129,'W',131-135;466-468,'X',470 ##label RO2
    ##note          this material was designated peptide M393 but may
                     contain two peptides
  #accession    C35483
    ##molecule_type protein
    ##residues      252-266,'H',268-272,'X',274-279,'I',281-284,'I',286
                     ##label RO3
    ##note          this proteolytic fragment was designated peptide M401
REFERENCE       A60432
  #authors      Zafar, R.S.; Walz, D.A.
  #journal      Thromb. Res. (1989) 53:31-44
  #title        Platelet membrane glycoprotein V: characterization of the
                  thrombin-sensitive glycoprotein from human platelets.
  #cross-references MUID:89162331
  #accession    A60432
    ##molecule_type protein
    ##residues      477-478,'FX',481-485,'E',487,'V',489-492,'NQ',495,'E',
                     497-498 ##label ZAF
REFERENCE       A47507
  #authors      Hickey, M.J.; Hagen, F.S.; Yagi, M.; Roth, G.J.
  #journal      Proc. Natl. Acad. Sci. U.S.A. (1993) 90:8327-8331
  #title        Human platelet glycoprotein V: characterization of the
                  polypeptide and the related Ib-V-IX receptor system of
                  adhesive, leucine-rich glycoproteins.
  #cross-references MUID:93391348
  #accession    A47507
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues      1-560 ##label RES
    ##cross-references GB:L11238; NID:g388759; PID:g388760
COMMENT         This platelet membrane protein is a substrate for thrombin.
COMMENT         The amino end of the intact protein is blocked.
COMMENT         This protein is absent in Bernard-Soulier syndrome.
GENETICS
  #gene         GDB:GP5
    ##cross-references GDB:230236; OMIM:173511
  #map_position 5pter-5qter
CLASSIFICATION  #superfamily leucine-rich alpha-2-glycoprotein repeat
                  homology
KEYWORDS        blocked amino end; glycoprotein; platelet; tandem repeat;
                  transmembrane protein
SUMMARY         #length 560  #molecular-weight 60958  #checksum 7673

  Query Match           26.6%;  Score 258;  DB 2;  Length 560;
  Best Local Similarity  32.8%;  Pred. No. 7.04e-19;
  Matches   43;  Conservative   37;  Mismatches  51;  Indels   0;  Gaps   0;

Db     76 LQRLMISDSHISAVAPGTFSDLIKLKTLRLSRNKITHLPGALLDKMVLLEQLFLDHNALR 135
          |: | : :::||::  |:| ||  |  |||:||::  :|  |:     | :| |  | ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    136 GIDQNMFQKLVNLQELALNQNQLDFLPASLFTNLENLKLLDLSGNNLTHLPKGLLGAQAK 195
          :| :: |:  |:: :| |: ||:  :  : |  | :| :| |: ||:|:|: : :   :|
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    196 LERLLLHSNRL 206
          |  : |||| |
Qy    122 LRTFRLHSNNL 132


RESULT   10
ENTRY           A41915     #type complete
TITLE           insulin-like growth factor-binding complex acid-labile chain
                  precursor - human
ALTERNATE_NAMES Acid-Labile Subunit (ALS)
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
                  20-Mar-1998
ACCESSIONS      A41915
REFERENCE       A41915
  #authors      Leong, S.R.; Baxter, R.C.; Camerato, T.; Dai, J.; Wood, W.I.
  #journal      Mol. Endocrinol. (1992) 6:870-876
  #title        Structure and functional expression of the acid-labile
                  subunit of the insulin-like growth factor-binding protein
                  complex.
  #cross-references MUID:92357025
  #accession    A41915
    ##status       preliminary
```

```
     ##molecule_type mRNA; protein
     ##residues      1-605 ##label LEO
     ##cross-references GB:M86826; NID:g184807; PID:g184808
     ##experimental_source liver
     ##note          sequence extracted from NCBI backbone (NCBIP:110171)
CLASSIFICATION   #superfamily leucine-rich alpha-2-glycoprotein repeat
                 homology
FEATURE
   75-98             #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   99-122            #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   123-146           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   147-170           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   171-194           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   195-218           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6\
   219-242           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR7\
   243-266           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR8\
   267-290           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR9\
   291-314           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR10\
   315-338           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR11\
   339-362           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR12\
   363-386           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR13\
   387-410           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR14\
   411-434           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR15\
   435-458           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR16\
   459-482           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR17\
   483-506           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR18\
   507-529           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR19
SUMMARY          #length 605  #molecular-weight 66034  #checksum 1870

  Query Match           26.0%;  Score 252;  DB 2;  Length 605;
  Best Local Similarity  32.8%;  Pred. No. 4.48e-18;
  Matches    43;  Conservative   30;  Mismatches  58;  Indels   0;  Gaps   0;

Db    220 LRELDLSRNALRAIKANVFVQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279
          || |:|  | : :|    | :|  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHNRIRQLAERSFEGLGQ 339
          ::   :| | : :: |:|  | |: :   :|: |: || | | :| | :|:  ||: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLDHNQL 350
          |  : |  | |
Qy    122 LRTFRLHSNNL 132


RESULT  11
ENTRY            JC5239     #type complete
TITLE            insulin-like growth factor acid-labile chain - baboon
ORGANISM         #formal_name Papio sp. #common_name baboon
DATE             17-Apr-1997 #sequence_revision 09-May-1997 #text_change
                   09-May-1997
ACCESSIONS       JC5239
REFERENCE        JC5239
   #authors      Delhanty, P.; Baxter, R.C.
   #journal      Biochem. Biophys. Res. Commun. (1996) 227:897-902
   #title        The cloning and expression of the baboon acid-labile subunit
                   of the insulin-like growth factor binding protein complex.
   #cross-references MUID:97040714
   #contents     liver
   #accession    JC5239
     ##molecule_type mRNA
     ##residues      1-605 ##label DEL
COMMENT      This factor is structurally related to proinsulin and have
               insuline-like metabolic, differentiative, and cell proliferative
               activities.
SUMMARY          #length 605  #molecular-weight 66110  #checksum 1703

  Query Match           25.5%;  Score 248;  DB 2;  Length 605;
  Best Local Similarity  32.8%;  Pred. No. 1.53e-17;
  Matches    43;  Conservative   29;  Mismatches  59;  Indels   0;  Gaps   0;

Db    220 LRELDLSRNALRAIKANVFAQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279
          || |:|  | : :|    | :|  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQ 339
          ::   :| | : :: |:|  | |: :   :|  |: || | | :| | :|:  ||: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLDHNQL 350
          |  : |  | |
Qy    122 LRTFRLHSNNL 132


RESULT  12
ENTRY            JC1282     #type complete
TITLE            insulin-like growth factor-binding protein acid labile chain
                   precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                   15-Aug-1997
ACCESSIONS       JC1282
REFERENCE        JC1282
   #authors      Dai, J.; Baxter, R.C.
   #journal      Biochem. Biophys. Res. Commun. (1992) 188:304-309
   #title        Molecular cloning of the acid-labile subunit of the rat
                   insulin-like growth factor binding protein complex.
   #cross-references MUID:93038676
   #accession    JC1282
     ##molecule_type mRNA
     ##residues      1-603 ##label DAI
     ##experimental_source liver
     ##note          the authors translated the codon AAG for residue 63 as
                      Arg, AAA for residue 205 as Pro and GGT for residue
                      260 as Arg
CLASSIFICATION   #superfamily leucine-rich alpha-2-glycoprotein repeat
                   homology
FEATURE
   1-27              #domain signal sequence #status predicted #label SIG\
   28-603            #product insulin-like growth factor binding protein,
                       acid labile chain #status predicted #label MAT
SUMMARY          #length 603  #molecular-weight 66811  #checksum 8075

  Query Match           24.9%;  Score 242;  DB 2;  Length 603;
  Best Local Similarity  34.1%;  Pred. No. 9.62e-17;
  Matches    46;  Conservative   34;  Mismatches  52;  Indels   3;  Gaps   3;

Db    240 HLPRLQKLYLDRNLITAVAPGAFLGMKALRWLDLSHNRVAGLMEDTFPGLLGLHVLRLAH 299
          ||  || | :: | |:::  |||  :| | :| |::| :  : |  | |   |  | |:
Qy      1 HLRVLQ-L-ME-NRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSE 57

Db    300 NAIASLRPRTFKDLHFLEELQLGHNRIRQLGERTFEGLGQLEVLTLNDNQITEVRVGAFS 359
          | | ::  ::|:    : :|||  |:|  : : :| :| :|||||||:| || : |::|:
Qy     58 NQIQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFN 117

Db    360 GLFNVAVMNLSGNCL 374
```

```
          : ::    | :| |
Qy    118 HMPKLRTFRLHSNNL 132


RESULT   13
ENTRY            JC6128     #type complete
TITLE            insulin-like growth factor binding complex acid labile chain
                   - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             23-Mar-1997 #sequence_revision 09-May-1997 #text_change
                   10-Sep-1997
ACCESSIONS       JC6128
REFERENCE        JC6128
   #authors      Boisclair, Y.R.; Seto, D.; Hsieh, S.; Hurst, K.R.; Ooi, G.T.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1996) 93:10028-10033
   #title        Organization and chromosomal localization of the gene
                   encoding the mouse acid labile subunit of the insulin-like
                   growth factor binding complex.
   #cross-references MUID:96413591
   #accession    JC6128
      ##molecule_type DNA
      ##residues      1-603 ##label BOI
      ##cross-references GB:U66900; NID:g1621612; PID:g1621613
COMMENT     This protein is a serum protein and it is of the ternary complex in
              the physiology of circulating insulin-like growth factor.
GENETICS
   #gene         als
   #map_position 17
SUMMARY          #length 603  #molecular-weight 66959  #checksum 7670

  Query Match           24.8%;  Score 241;  DB 2;  Length 603;
  Best Local Similarity  30.5%;  Pred. No. 1.31e-16;
  Matches    40;  Conservative   32;  Mismatches  59;  Indels   0;  Gaps   0;

Db    220 LRELDLSRNALRSVKANVFIHLPRLQKLYLDRNLITAVAPRAFLGMKALRWLDLSHNRVA 279
          || |:|  | : ::    |  |  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLHVLRLAHNAITSLRPRTFKDLHFLEELQLGHNRIRQLGEKTFEGLGQ 339
          ::   :| | : :  |:|  | |: :   :|: |: || | | :| | :|:  :|: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLNDNQI 350
          |  : |: | :
Qy    122 LRTFRLHSNNL 132


RESULT   14
ENTRY            S40241     #type complete
TITLE            G protein-coupled receptor - great pond snail
ORGANISM         #formal_name Lymnaea stagnalis #common_name great pond snail
DATE             06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                   09-Sep-1997
ACCESSIONS       S40241
REFERENCE        S40241
   #authors      Tensen, C.P.; Kesteren, E.R.; Planta, R.J.; Cox, K.; Burke,
                   J.F.; Heerikhuizen, H.; Vreugdenhil, E.
   #submission   submitted to the EMBL Data Library, June 1993
   #description  A G protein-coupled receptor with LDL-binding motifs suggests
                   a role for lipoproteins in G-linked signal transduction.
   #accession    S40241
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-1115 ##label TEN
      ##cross-references EMBL:Z23104; NID:g438128; PID:g438129
CLASSIFICATION   #superfamily LDL receptor ligand-binding repeat homology;
                   leucine-rich alpha-2-glycoprotein repeat homology
KEYWORDS         G protein-coupled receptor; transmembrane protein
FEATURE
   38-77             #domain LDL receptor ligand-binding repeat homology
                       #label LDL1\
   79-113            #domain LDL receptor ligand-binding repeat homology
                       #label LDL2\
   158-194           #domain LDL receptor ligand-binding repeat homology
                       #label LDL3\
   233-267           #domain LDL receptor ligand-binding repeat homology
                       #label LDL4\
   322-361           #domain LDL receptor ligand-binding repeat homology
                       #label LDL5\
   367-401           #domain LDL receptor ligand-binding repeat homology
                       #label LDL6\
   446-483           #domain LDL receptor ligand-binding repeat homology
                       #label LDL7\
   488-523           #domain LDL receptor ligand-binding repeat homology
                       #label LDL8\
   584-607           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   608-631           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   632-655           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   656-679           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   704-727           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   774-797           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6
SUMMARY          #length 1115  #molecular-weight 125864  #checksum 74

  Query Match           24.4%;  Score 237;  DB 2;  Length 1115;
  Best Local Similarity  28.6%;  Pred. No. 4.41e-16;
  Matches    38;  Conservative   40;  Mismatches  55;  Indels   0;  Gaps   0;

Db    609 LTHLNLADNNITSLKNGSLLGLSNLKQLHINGNKIETIEEDTFSSMIHLTVLDLSNQRLT 668
          |  | | :| |:::  |::  | :| :|::| |::: : |  | :  :|  ||||: ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    669 HVYKNMFKGLKQITVLNISRNQINSIDNGAFNNLANVRLIDLSGNVIKDIGQKVFMGLPR 728
          : :: |:|  :|  | :  |||: |::||| | :: :: |: | |  ::   |  :|:
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    729 LVELKTDSYRFCC 741
          |  ::  |  : |
Qy    122 LRTFRLHSNNLYC 134


RESULT   15
ENTRY            NBHUA2     #type complete
TITLE            leucine-rich alpha-2-glycoprotein - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change
                   05-Dec-1998
ACCESSIONS       A03211
REFERENCE        A03211
   #authors      Takahashi, N.; Takahashi, Y.; Putnam, F.W.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1985) 82:1906-1910
   #title        Periodicity of leucine and tandem repetition of a 24-amino
                   acid segment in the primary structure of leucine-rich
                   alpha-2-glycoprotein of human serum.
   #cross-references MUID:85166241
   #accession    A03211
      ##molecule_type protein
      ##residues      1-312 ##label TAK
COMMENT     The function of this plasma protein is not known.
CLASSIFICATION   #superfamily leucine-rich alpha-2-glycoprotein; leucine-rich
                   alpha-2-glycoprotein repeat homology; proteoglycan
                   carboxyl-terminal homology
KEYWORDS         duplication; glycoprotein; plasma; tandem repeat
FEATURE
   58-81             #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   82-105            #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   106-129           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   130-153           #domain leucine-rich alpha-2-glycoprotein repeat
```

homology #label LRR4\
154-177 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR5\
178-201 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR6\
202-225 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR7\
226-249 #domain leucine-rich alpha-2-glycoprotein repeat homology #label LRR8\
262-309 #domain proteoglycan carboxyl-terminal homology #label PCH\
2 #binding_site carbohydrate (Thr) (covalent) #status experimental\
8-21,268-294 #disulfide_bonds #status experimental\
44,151,234,290 #binding_site carbohydrate (Asn) (covalent) #status experimental\
271 #binding_site carbohydrate (Asn) (covalent) #status absent
MARY #length 312 #molecular-weight 34346 #checksum 6045

Query Match 24.2%; Score 235; DB 1; Length 312;
Best Local Similarity 31.8%; Pred. No. 8.10e-16;
Matches 42; Conservative 39; Mismatches 49; Indels 2; Gaps 2;

```
Db    107 LDTLVLKENQLEVLEVSWLHGLKALGHLDLSGNRLRKLPPGL-LANFTLLRTLDLGENQL 165
          |  | |:||::  :| : :: || | :| |: | |: :|  | |:   | | |||:|||:
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYR-LDLSENQI 60

Db    166 ETLPPDLLRGPLQLERLHLEGNKLQVLGKDLLLPQPDLRYLFLNGNKLARVAAGAFQGLR 225
          :::|   :||::::  |:|: | :  :    : :  ||  | || |:::|:: ::|  :
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    226 QLDMLDLSNNSL 237
          |  : | :|:|
Qy    121 KLRTFRLHSNNL 132
```

Search completed: Fri May 28 09:30:18 1999
Job time : 21 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:30:36 1999; MasPar time 5.87 Seconds
645.690 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-12
Description: (1-134) from US09191647.pep
Perfect Score: 971
Sequence: 1 HLRVLQLMENRISTIERGAF.....SFNHMPKLRTFRLHSNNLYC 134

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 46.321; Variance 107.872; scale 0.429

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	366	37.7	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	1.73e-39
2	298	30.7	682	1	CONN_DROME	CONNECTIN PRECURSOR.	1.77e-28
3	267	27.5	567	1	GPV_RAT	PLATELET GLYCOPROTEIN	1.37e-23
4	264	27.2	1134	1	CHAO_DROME	CHAOPTIN PRECURSOR (PH	4.01e-23
5	260	26.8	536	1	CBP8_HUMAN	CARBOXYPEPTIDASE N 83	1.68e-22
6	259	26.7	567	1	GPV_MOUSE	PLATELET GLYCOPROTEIN	2.40e-22
7	258	26.6	560	1	GPV_HUMAN	PLATELET GLYCOPROTEIN	3.43e-22
8	252	26.0	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	2.91e-21
9	248	25.5	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	1.20e-20
10	242	24.9	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	1.00e-19
11	241	24.8	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	1.42e-19
12	237	24.4	1115	1	GPCR_LYMST	G-PROTEIN COUPLED RECE	5.79e-19
13	235	24.2	312	1	A2GL_HUMAN	LEUCINE-RICH ALPHA-2-G	1.17e-18
14	222	22.9	361	1	CHAD_BOVIN	CHONDROADHERIN PRECURS	1.07e-16
15	222	22.9	662	1	GARP_HUMAN	GARP PROTEIN PRECURSOR	1.07e-16
16	213	21.9	368	1	PGS1_HUMAN	BONE/CARTILAGE PROTEOG	2.36e-15
17	213	21.9	369	1	PGS1_MOUSE	BONE/CARTILAGE PROTEOG	2.36e-15
18	213	21.9	369	1	PGS1_RAT	BONE/CARTILAGE PROTEOG	2.36e-15
19	210	21.6	369	1	PGS1_CANFA	BONE/CARTILAGE PROTEOG	6.58e-15
20	210	21.6	369	1	PGS1_BOVIN	BONE/CARTILAGE PROTEOG	6.58e-15
21	208	21.4	1097	1	TOLL_DROME	TOLL PROTEIN PRECURSOR	1.30e-14
22	205	21.1	357	1	PGS2_CHICK	BONE PROTEOGLYCAN II P	3.60e-14
23	202	20.8	360	1	PGS2_BOVIN	BONE PROTEOGLYCAN II P	9.91e-14
24	196	20.2	359	1	PGS2_HUMAN	BONE PROTEOGLYCAN II P	7.43e-13
25	196	20.2	360	1	PGS2_CANFA	BONE PROTEOGLYCAN II P	7.43e-13
26	190	19.6	354	1	PGS2_MOUSE	BONE PROTEOGLYCAN II P	5.48e-12
27	189	19.5	382	1	PARG_HUMAN	PROLARGIN PRECURSOR (P	7.64e-12
28	188	19.4	925	1	GLHR_ANTEL	PROBABLE GLYCOPROTEIN	1.06e-11
29	182	18.7	1257	1	FLIH_CAEEL	FLIGHTLESS-I PROTEIN H	7.65e-11
30	180	18.5	354	1	PGS2_RAT	BONE PROTEOGLYCAN II P	1.47e-10
31	180	18.5	376	1	FMOD_MOUSE	FIBROMODULIN PRECURSOR	1.47e-10
32	180	18.5	376	1	FMOD_RAT	FIBROMODULIN PRECURSOR	1.47e-10
33	180	18.5	626	1	GPBA_HUMAN	PLATELET GLYCOPROTEIN	1.47e-10
34	179	18.4	277	1	RSU1_MOUSE	RAS SUPPRESSOR PROTEIN	2.04e-10
35	178	18.3	277	1	RSU1_HUMAN	RAS SUPPRESSOR PROTEIN	2.82e-10
36	175	18.0	1839	1	CYAA_SACKL	ADENYLATE CYCLASE (EC	7.46e-10
37	171	17.6	376	1	FMOD_HUMAN	FIBROMODULIN PRECURSOR	2.70e-09
38	171	17.6	2145	1	CYAA_PODAN	ADENYLATE CYCLASE (EC	2.70e-09
39	165	17.0	375	1	FMOD_BOVIN	FIBROMODULIN PRECURSOR	1.83e-08
40	164	16.9	701	1	LSHR_BOVIN	LUTROPIN-CHORIOGONADOT	2.52e-08
41	162	16.7	676	1	LSHR_CALJA	LUTROPIN-CHORIOGONADOT	4.73e-08
42	162	16.7	764	1	TSHR_SHEEP	THYROTROPIN RECEPTOR P	4.73e-08
43	160	16.5	763	1	TSHR_BOVIN	THYROTROPIN RECEPTOR P	8.87e-08
44	159	16.4	2026	1	CYAA_YEAST	ADENYLATE CYCLASE (EC	1.21e-07
45	157	16.2	1692	1	CYAA_SCHPO	ADENYLATE CYCLASE (EC	2.27e-07

ALIGNMENTS

```
RESULT   1
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1     36
FT   CHAIN       37   1480       SLIT PROTEIN.
FT   DOMAIN      70    104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105    230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231    294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295    326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327    452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453    518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519    550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551    653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654    714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715    746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747    848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849    910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105    115       LRR 1-1.
FT   REPEAT     116    139       LRR 1-2.
FT   REPEAT     140    163       LRR 1-3.
FT   REPEAT     164    187       LRR 1-4.
FT   REPEAT     188    211       LRR 1-5.
FT   REPEAT     212    230       LRR 1-6.
FT   REPEAT     327    337       LRR 2-1.
FT   REPEAT     338    361       LRR 2-2.
FT   REPEAT     362    385       LRR 2-3.
FT   REPEAT     386    409       LRR 2-4.
FT   REPEAT     410    433       LRR 2-5.
FT   REPEAT     434    452       LRR 2-6.
FT   REPEAT     551    562       LRR 3-1.
FT   REPEAT     563    586       LRR 3-2.
FT   REPEAT     587    610       LRR 3-3.
FT   REPEAT     611    634       LRR 3-4.
FT   REPEAT     635    653       LRR 3-5.
FT   REPEAT     747    757       LRR 4-1.
FT   REPEAT     758    781       LRR 4-2.
FT   REPEAT     782    805       LRR 4-3.
FT   REPEAT     806    829       LRR 4-4.
FT   REPEAT     830    848       LRR 4-5.
FT   DOMAIN     907    944       EGF-LIKE 1.
FT   DOMAIN     946    983       EGF-LIKE 2.
FT   DOMAIN     985   1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024   1062       EGF-LIKE 4.
FT   DOMAIN    1064   1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111   1149       EGF-LIKE 6.
FT   DOMAIN    1353   1392       EGF-LIKE 7.
FT   DOMAIN    1409   1480       CTCK.
FT   VARSPLIC  1394   1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111    111       POTENTIAL.
FT   CARBOHYD   207    207       POTENTIAL.
FT   CARBOHYD   357    357       POTENTIAL.
FT   CARBOHYD   435    435       POTENTIAL.
FT   CARBOHYD   783    783       POTENTIAL.
FT   CARBOHYD   788    788       POTENTIAL.
FT   CARBOHYD   958    958       POTENTIAL.
FT   CARBOHYD   998    998       POTENTIAL.
FT   CARBOHYD  1060   1060       POTENTIAL.
FT   CARBOHYD  1159   1159       POTENTIAL.
FT   CARBOHYD  1175   1175       POTENTIAL.
FT   CARBOHYD  1243   1243       POTENTIAL.
FT   CARBOHYD  1292   1292       POTENTIAL.
FT   DISULFID   911    922       BY SIMILARITY.
FT   DISULFID   916    932       BY SIMILARITY.
FT   DISULFID   934    943       BY SIMILARITY.
FT   DISULFID   950    961       BY SIMILARITY.
FT   DISULFID   955    971       BY SIMILARITY.
FT   DISULFID   973    982       BY SIMILARITY.
FT   DISULFID   989   1001       BY SIMILARITY.
FT   DISULFID   995   1010       BY SIMILARITY.
FT   DISULFID  1012   1021       BY SIMILARITY.
FT   DISULFID  1028   1041       BY SIMILARITY.
FT   DISULFID  1035   1050       BY SIMILARITY.
FT   DISULFID  1052   1061       BY SIMILARITY.
FT   DISULFID  1068   1079       BY SIMILARITY.
FT   DISULFID  1073   1088       BY SIMILARITY.
FT   DISULFID  1090   1099       BY SIMILARITY.
FT   DISULFID  1115   1125       BY SIMILARITY.
FT   DISULFID  1120   1137       BY SIMILARITY.
FT   DISULFID  1139   1148       BY SIMILARITY.
FT   DISULFID  1357   1368       BY SIMILARITY.
FT   DISULFID  1362   1380       BY SIMILARITY.
FT   DISULFID  1382   1391       BY SIMILARITY.
FT   DISULFID  1409   1443       BY SIMILARITY.
FT   DISULFID  1423   1457       BY SIMILARITY.
FT   DISULFID  1434   1473       BY SIMILARITY.
FT   DISULFID  1438   1475       BY SIMILARITY.
FT   DISULFID  1442   1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          37.7%;  Score 366;  DB 1;  Length 1480;
  Best Local Similarity  40.0%;  Pred. No. 1.73e-39;
  Matches   52;  Conservative   31;  Mismatches  47;  Indels   0;  Gaps   0;

Db    105 LELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVITTVG 164
          |:|  | :: |    || |  |  |:|  | :: :    |     | |||:|:| | ::
Qy      5 LQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQAIP 64

Db    165 RRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGGLGRLRA 224
          |: |:||  :::|||| |||:|::: ||::| :||:||||||:| |:   |  : :||:
Qy     65 RKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRT 124

Db    225 LRLSDNPFAC 234
          :||  | : |
Qy    125 FRLHSNNLYC 134


RESULT    2
ID   CONN_DROME     STANDARD;      PRT;   682 AA.
AC   Q01819;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CONNECTIN PRECURSOR.
GN   CON.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92370678.
RA   NOSE A., MAHAJAN V.B., GOODMAN C.S.;
RT   "Connectin: a homophilic cell adhesion molecule expressed on a subset
RT   of muscles and the motoneurons that innervate them in Drosophila.";
RL   CELL 70:553-567(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 93202002.
RA   GOULD A.P., WHITE R.A.H.;
RT   "Connectin, a target of homeotic gene control in Drosophila.";
RL   DEVELOPMENT 116:1163-1174(1992).
CC   -!- FUNCTION: CELL ADHESION PROTEIN INVOLVED IN TARGET RECOGNITION
CC       DURING NEUROMUSCULAR DEVELOPMENT. MEDIATES HOMOPHILIC CELLULAR
CC       ADHESION.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN ABDOMINAL AND
CC       THORACIC SEGMENT MUSCLE AND MOTORNEURON CELLS.
CC   -!- DEVELOPMENTAL STAGE: EMBRYO.
```

```
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN:10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96647; G157084; -.
DR   EMBL; X68701; G7738; -.
DR   PIR; S28464; S28464.
DR   PIR; A43318; A43318.
DR   FLYBASE; FBgn0005775; Con.
DR   PFAM; PF00560; LRR; 5.
KW   CELL ADHESION; DEVELOPMENTAL PROTEIN; EMBRYO; SIGNAL; GPI-ANCHOR;
KW   LEUCINE-REPEAT; REPEAT.
FT   SIGNAL       1    24
FT   CHAIN       25   665       CONNECTIN.
FT   PROPEP     666   682       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN     142   381       LEUCINE-RICH REPEATS.
FT   REPEAT     142   165       LRR 1.
FT   REPEAT     166   189       LRR 2.
FT   REPEAT     190   213       LRR 3.
FT   REPEAT     214   237       LRR 4.
FT   REPEAT     238   261       LRR 5.
FT   REPEAT     262   285       LRR 6.
FT   REPEAT     286   299       LRR 7.
FT   REPEAT     300   322       LRR 8.
FT   REPEAT     324   347       LRR 9.
FT   REPEAT     348   381       LRR 10.
FT   LIPID      665   665       GPI-ANCHOR (POTENTIAL).
FT   CONFLICT   631   631       E -> G (IN REF. 2).
FT   CONFLICT   674   677       QVAL -> VALM (IN REF. 2).
SQ   SEQUENCE   682 AA;  75992 MW;  3E15592A CRC32;

  Query Match           30.7%;  Score 298;  DB 1;  Length 682;
  Best Local Similarity  38.3%;  Pred. No. 1.77e-28;
  Matches   51;  Conservative   28;  Mismatches  52;  Indels   2;  Gaps   2;

Db    223 RLRELNLEHNQIFEMDRYAFRNL-PLCERLFLNNNNISTLHEGLFADMARLTFLNLAHNQ 281
          :|| | |  |:|  ::| ||::|  | ||| || ||:  : | ||   |||  |:|: ||
Qy      1 HLRVLQLMENRISTIERGAFQDLKEL-ERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db    282 INVLTSEIFRGLGNLNVLKLTRNNLNFIGDTVFAELWSLSELELDDNRIERISERALDGL 341
          |  :    |||  ::: | |  | :: | |  |  |: |  | |::| | |:|  ::: :
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    342 NTLKTLNLRNNLL 354
            |:|: |::| |
Qy    120 PKLRTFRLHSNNL 132


RESULT   3
ID   GPV_RAT       STANDARD;      PRT;   567 AA.
AC   O08770;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RX   MEDLINE; 97275136.
RA   RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,
RA   GRUNERT P., LOEW D., VAN DORSSELAER A., CAZENAVE J.-P., LANZA F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   BLOOD 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69594; E222201; -.
DR   PFAM; PF00560; LRR; 8.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17   522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   523   543       POTENTIAL.
FT   DOMAIN     544   567       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55   415       LEUCINE-RICH REPEATS.
FT   REPEAT      55    78       LRR 1.
FT   REPEAT      79   102       LRR 2.
FT   REPEAT     103   126       LRR 3.
FT   REPEAT     127   150       LRR 4.
FT   REPEAT     151   174       LRR 5.
FT   REPEAT     175   198       LRR 6.
FT   REPEAT     199   222       LRR 7.
FT   REPEAT     223   246       LRR 8.
FT   REPEAT     247   270       LRR 9.
FT   REPEAT     271   294       LRR 10.
FT   REPEAT     295   318       LRR 11.
FT   REPEAT     319   343       LRR 12.
FT   REPEAT     346   367       LRR 13.
FT   REPEAT     368   391       LRR 14.
FT   REPEAT     392   415       LRR 15.
FT   CARBOHYD    51    51       POTENTIAL.
FT   CARBOHYD   181   181       POTENTIAL.
FT   CARBOHYD   243   243       POTENTIAL.
FT   CARBOHYD   298   298       POTENTIAL.
FT   CARBOHYD   312   312       POTENTIAL.
FT   CARBOHYD   385   385       POTENTIAL.
FT   CARBOHYD   498   498       POTENTIAL.
SQ   SEQUENCE   567 AA;  63344 MW;  ABAEC91D CRC32;

  Query Match           27.5%;  Score 267;  DB 1;  Length 567;
  Best Local Similarity  32.6%;  Pred. No. 1.37e-23;
  Matches   43;  Conservative   32;  Mismatches  56;  Indels   1;  Gaps   1;

Db    220 LTELRLERNHLRSIAPGAFDSLGNLSTLTLSGNLLESLPPALFLHVSWLTRLTLFENPLE 279
          |  |:|  |:: :|  |||: | :|  | |: | |: :|  |||  ::| || | || ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 ELPEVLFGEMAGLRELWLNGTHLRTLPAAAFRNLSGLQTLGLTRNPLLSALPPGMFHGLT 339
            :|   |    :::| |:  ::  :  :|| |  |: | |  | : : |: : |: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNI-TRLSVASFNHMP 120

Db    340 ELRVLAVHTNAL 351
          || : :|:| |
Qy    121 KLRTFRLHSNNL 132


RESULT   4
ID   CHAO_DROME     STANDARD;      PRT;  1134 AA.
```

```
AC   P12024;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CHAOPTIN PRECURSOR (PHOTORECEPTOR CELL-SPECIFIC MEMBRANE PROTEIN).
GN   CHP OR CHT.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88135762.
RA   REINKE R., KRANTZ D.E., YEN D., ZIPURSKY S.L.;
RT   "Chaoptin, a cell surface glycoprotein required for Drosophila
RT   photoreceptor cell morphogenesis, contains a repeat motif found in
RT   yeast and human.";
RL   CELL 52:291-301(1988).
CC   -!- FUNCTION: REQUIRED FOR DROSOPHILA PHOTORECEPTOR CELL
CC       MORPHOGENESIS. MEDIATES HOMOPHILIC CELLULAR ADHESION.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR SURFACE OF R-CELL PLASMA
CC       MEMBRANE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED 24 HOURS AFTER INITIATION OF
CC       PHOTORECEPTOR CELL DIFFERENTIATION, PERSISTS THROUGH ADULTHOOD.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 41.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19017; G157098; -.
DR   EMBL; M19008; G157098; JOINED.
DR   EMBL; M19009; G157098; JOINED.
DR   EMBL; M19010; G157098; JOINED.
DR   EMBL; M19011; G157098; JOINED.
DR   EMBL; M19012; G157098; JOINED.
DR   EMBL; M19013; G157098; JOINED.
DR   EMBL; M19014; G157098; JOINED.
DR   EMBL; M19016; G157098; JOINED.
DR   PIR; A29944; A29944.
DR   FLYBASE; FBgn0000313; chp.
DR   PFAM; PF00560; LRR; 17.
KW   GLYCOPROTEIN; MEMBRANE; SIGNAL; REPEAT; LEUCINE-REPEAT; VISION.
FT   SIGNAL       1     29
FT   CHAIN       30   1134       CHAOPTIN.
FT   CARBOHYD    77     77       POTENTIAL.
FT   CARBOHYD   267    267       POTENTIAL.
FT   CARBOHYD   305    305       POTENTIAL.
FT   CARBOHYD   339    339       POTENTIAL.
FT   CARBOHYD   361    361       POTENTIAL.
FT   CARBOHYD   422    422       POTENTIAL.
FT   CARBOHYD   680    680       POTENTIAL.
FT   CARBOHYD   692    692       POTENTIAL.
FT   CARBOHYD   718    718       POTENTIAL.
FT   CARBOHYD   746    746       POTENTIAL.
FT   CARBOHYD   936    936       POTENTIAL.
FT   CARBOHYD   970    970       POTENTIAL.
FT   CARBOHYD  1012   1012       POTENTIAL.
FT   CARBOHYD  1104   1104       POTENTIAL.
SQ   SEQUENCE   1134 AA;  130719 MW;  B67A6363`CRC32;

  Query Match          27.2%;  Score 264;  DB 1;  Length 1134;
  Best Local Similarity  35.6%;  Pred. No. 4.01e-23;
  Matches   48;  Conservative   35;  Mismatches  47;  Indels   5;  Gaps   5;

Db     80 KVFMLHMENTGLREIEP-YFLQSTGMYRLKISGNHLTEIPDDAFTGLERSLWELILPQND 138
          :|: | |||  :  ||   | :   : ||::: |:|  :|:  | |  | |  | |::|:
Qy      3 RVLQL-MENR-ISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTAR-LYRLDLSENQ 59

Db    139 LVEIPSKSLRHLQKLRHLDLGYNHITHIQHDSFRGLEDSLQTLILRENCISQLMSHSFSG 198
          :  || |::|    :::|:| ||:|: |:  :||:| | |: | | :| |::|   ||:
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRD-LEVLTLNNNNITRLSVASFNH 118

Db    199 LLILETLDLSGNNLF 213
          :  | |: | :|||:
Qy    119 MPKLRTFRLHSNNLY 133


RESULT    5
ID   CBP8_HUMAN     STANDARD;      PRT;   536 AA.
AC   P22792;
DT   01-AUG-1991 (REL. 19, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CARBOXYPEPTIDASE N 83 KD CHAIN (CARBOXYPEPTIDASE N REGULATORY
DE   SUBUNIT) (FRAGMENT).
GN   CPN2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 90094386.
RA   TAN F., WEERASINGHE D.K., SKIDGEL R.A., TAMEI H., KAUL R.K.,
RA   RONINSON I.B., SCHILLING J.W., ERDOES E.G.;
RT   "The deduced protein sequence of the human carboxypeptidase N high
RT   molecular weight subunit reveals the presence of leucine-rich tandem
RT   repeats.";
RL   J. BIOL. CHEM. 265:13-19(1990).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE; 88309120.
RA   SKIDGEL R.A., BENNETT C.D., SCHILLING J.W., TAN F., WEERASINGHE D.K.,
RA   ERDOES E.G.;
RT   "Amino acid sequence of the N-terminus and selected tryptic peptides
RT   of the active subunit of human plasma carboxypeptidase N: comparison
RT   with other carboxypeptidases.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 154:1323-1329(1988).
CC   -!- FUNCTION: THE 83 KD SUBUNIT BINDS AND STABILIZES THE CATALYTIC
CC       SUBUNIT AT 37 DEGREES CELSIUS AND KEEPS IT IN CIRCULATION. UNDER
CC       SOME CIRCUMSTANCES IT MAY BE AN ALLOSTERIC MODIFIER OF THE
CC       CATALYTIC SUBUNIT.
CC   -!- SUBUNIT: TETRAMER OF TWO CATALYTIC CHAINS AND TWO GLYCOSYLATED
CC       INACTIVE CHAINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGION (POTENTIAL).
CC   -!- PTM: WHETHER OR NOT ANY CYS RESIDUES PARTICIPATE IN INTRACHAIN
CC       BONDS IS UNKNOWN, BUT THEY DO NOT FORM INTERCHAIN DISULFIDE BONDS
CC       WITH THE 50 KD CATALYTIC SUBUNIT.
CC   -!- DISEASE: A COMPLETE ABSENCE OF THE ENZYME IS NOT CONSIDERED TO BE
CC       COMPATIBLE WITH LIFE.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 12.
CC   -!- SIMILARITY: SOME, TO E.COLI YDDK.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05158; G179936; -.
DR   PIR; A34901; A34901.
DR   MIM; 603104; -.
DR   PFAM; PF00560; LRR; 7.
KW   REPEAT; LEUCINE-REPEAT; GLYCOPROTEIN.
FT   NON_TER      1      1
FT   DOMAIN      68    355       LEUCINE-RICH REPEATS.
```

```
FT   REPEAT       68     91       LRR 1.
FT   REPEAT       92    115       LRR 2.
FT   REPEAT      116    139       LRR 3.
FT   REPEAT      140    163       LRR 4.
FT   REPEAT      164    187       LRR 5.
FT   REPEAT      188    211       LRR 6.
FT   REPEAT      212    235       LRR 7.
FT   REPEAT      236    259       LRR 8.
FT   REPEAT      260    283       LRR 9.
FT   REPEAT      284    307       LRR 10.
FT   REPEAT      308    331       LRR 11.
FT   REPEAT      332    355       LRR 12.
FT   DOMAIN      359    379       THR/SER-RICH.
FT   CARBOHYD     53     53       POTENTIAL.
FT   CARBOHYD     90     90       POTENTIAL.
FT   CARBOHYD     98     98       POTENTIAL.
FT   CARBOHYD    207    207       POTENTIAL.
FT   CARBOHYD    245    245       POTENTIAL.
FT   CARBOHYD    327    327       POTENTIAL.
FT   CARBOHYD    338    338       POTENTIAL.
FT   CARBOHYD    495    495       POTENTIAL.
SQ   SEQUENCE   536 AA;  58649 MW;  C4413E03 CRC32;

  Query Match           26.8%;  Score 260;  DB 1;  Length 536;
  Best Local Similarity  32.6%;  Pred. No. 1.68e-22;
  Matches   43;  Conservative   31;  Mismatches  58;  Indels   0;  Gaps   0;

Db     77 RLEDLEVTGSSFLNLSTNIFSNLTSLGKLTLNFNMLEALPEGLFQHLAALESLHLQGNQL 136
          :|  |::  : :  :    | :|  | :| || | |: :|| ||   | |  | |  ||:
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    137 QALPRRLFQPLTHLKTLNLAQNLLAQLPEELFHPLTSLQTLKLSNNALSGLPQGVFGKLG 196
          ||:||: |:    :| | |  | :: : :  |::|  |: | |:|| :: |: : |  :
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    197 SLQELFLDSNNI 208
            |: : | |||:
Qy    121 KLRTFRLHSNNL 132


RESULT    6
ID   GPV_MOUSE      STANDARD;      PRT;   567 AA.
AC   O08742;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=LIVER;
RX   MEDLINE; 97275136.
RA   RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,
RA   GRUNERT P., LOEW D., VAN DORSSELAER A., CAZENAVE J.-P., LANZA F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   BLOOD 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69595; E222202; -.
DR   MGD; MGI:1096363; GP5.
DR   PFAM; PF00560; LRR; 7.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1     16       POTENTIAL.
FT   CHAIN       17    567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   523    543       POTENTIAL.
FT   DOMAIN     544    567       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55    415       LEUCINE-RICH REPEATS.
FT   REPEAT      55     78       LRR 1.
FT   REPEAT      79    102       LRR 2.
FT   REPEAT     103    126       LRR 3.
FT   REPEAT     127    150       LRR 4.
FT   REPEAT     151    174       LRR 5.
FT   REPEAT     175    198       LRR 6.
FT   REPEAT     199    222       LRR 7.
FT   REPEAT     223    246       LRR 8.
FT   REPEAT     247    270       LRR 9.
FT   REPEAT     271    294       LRR 10.
FT   REPEAT     295    318       LRR 11.
FT   REPEAT     319    343       LRR 12.
FT   REPEAT     346    367       LRR 13.
FT   REPEAT     368    391       LRR 14.
FT   REPEAT     392    415       LRR 15.
FT   CARBOHYD    51     51       POTENTIAL.
FT   CARBOHYD    67     67       POTENTIAL.
FT   CARBOHYD   181    181       POTENTIAL.
FT   CARBOHYD   243    243       POTENTIAL.
FT   CARBOHYD   298    298       POTENTIAL.
FT   CARBOHYD   312    312       POTENTIAL.
FT   CARBOHYD   385    385       POTENTIAL.
SQ   SEQUENCE   567 AA;  63467 MW;  3AE7515E CRC32;

  Query Match           26.7%;  Score 259;  DB 1;  Length 567;
  Best Local Similarity  32.6%;  Pred. No. 2.40e-22;
  Matches   43;  Conservative   31;  Mismatches  57;  Indels   1;  Gaps   1;

Db    220 LTELRLERNHLRSVAPGAFDRLGNLSSLTLSGNLLESLPPALFLHVSSVSRLTLFENPLE 279
          | |:|  |:: ::  |||: | :|  | |: | |: :|  |||  : : || | || ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 ELPDVLFGEMAGLRELWLNGTHLSTLPAAAFRNLSGLQTLGLTRNPRLSALPRGVFQGLR 339
          :|   |     :::| |: ::| :  :||| |  |: | |  |  :: |: : |  :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNN-NITRLSVASFNHMP 120

Db    340 ELRVLGLHTNAL 351
            || : ||:| |
Qy    121 KLRTFRLHSNNL 132


RESULT    7
ID   GPV_HUMAN      STANDARD;      PRT;   560 AA.
AC   P40197;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN   GP5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 93391348.
RA   HICKEY M.J., HAGEN F.S., YAGI M., ROTH G.J.;
```

```
RT   "Human platelet glycoprotein V: characterization of the polypeptide
RT   and the related Ib-V-IX receptor system of adhesive, leucine-rich
RT   glycoproteins.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:8327-8331(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLATELET;
RX   MEDLINE; 94012616.
RA   LANZA F., MORALES M., DE LA SALLE C., CAZENAVE J.-P., CLEMETSON K.J.,
RA   SHIMOMURA T., PHILLIPS D.R.;
RT   "Cloning and characterization of the gene encoding the human platelet
RT   glycoprotein V. A member of the leucine-rich glycoprotein family
RT   cleaved during thrombin-induced platelet activation.";
RL   J. BIOL. CHEM. 268:20801-20807(1993).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   TISSUE=PLATELET;
RX   MEDLINE; 90275263.
RA   SHIMOMURA T., FUJIMURA K., MAEHAMA S., TAKEMOTO M., ODA K.,
RA   FUJIMOTO T., OYAMA R., SUZUKI M., ICIHARA-TANAKA K., TITANI K.,
RA   KURAMOTO A.;
RT   "Rapid purification and characterization of human platelet
RT   glycoprotein V: the amino acid sequence contains leucine-rich
RT   repetitive modules as in glycoprotein Ib.";
RL   BLOOD 75:2349-2356(1990).
RN   [4]
RP   PARTIAL SEQUENCE.
RC   TISSUE=PLATELET;
RX   MEDLINE; 90321220.
RA   ROTH G.J., CHURCH T.A., MCMULLEN B.A., WILLIAMS S.A.;
RT   "Human platelet glycoprotein V: a surface leucine-rich glycoprotein
RT   related to adhesion.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 170:153-161(1990).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: PLATELETS AND MEGAKARYOCYTES.
CC   -!- PTM: THE N-TERMINAL IS BLOCKED.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L11238; G388760; -.
DR   EMBL; Z23091; G312502; -.
DR   MIM; 173511; -.
DR   PFAM; PF00560; LRR; 8.
DR   HSSP; P16473; 1XUM.
KW   PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;
KW   REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   560       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17   523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   524   544       POTENTIAL.
FT   DOMAIN     545   560       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      55   415       LEUCINE-RICH REPEATS.
FT   REPEAT      55    78       LRR 1.
FT   REPEAT      79   102       LRR 2.
FT   REPEAT     103   126       LRR 3.
FT   REPEAT     127   150       LRR 4.
FT   REPEAT     151   174       LRR 5.
FT   REPEAT     175   198       LRR 6.
FT   REPEAT     199   222       LRR 7.
FT   REPEAT     223   246       LRR 8.
FT   REPEAT     247   270       LRR 9.
FT   REPEAT     271   294       LRR 10.
FT   REPEAT     295   318       LRR 11.
FT   REPEAT     319   343       LRR 12.
FT   REPEAT     346   367       LRR 13.
FT   REPEAT     368   391       LRR 14.
FT   REPEAT     392   415       LRR 15.
FT   CARBOHYD    51    51
FT   CARBOHYD   181   181
FT   CARBOHYD   243   243       POTENTIAL.
FT   CARBOHYD   267   267       POTENTIAL.
FT   CARBOHYD   298   298
FT   CARBOHYD   312   312
FT   CARBOHYD   385   385
FT   CARBOHYD   499   499       POTENTIAL.
FT   CONFLICT    73    74       MT -> TK (IN REF. 2).
FT   CONFLICT   109   109       K -> T (IN REF. 2).
FT   CONFLICT   130   130       D -> W (IN REF. 3).
FT   CONFLICT   136   138       GID -> PGG (IN REF. 3).
FT   CONFLICT   209   209       L -> I (IN REF. 2).
FT   CONFLICT   267   267       N -> H (IN REF. 3).
FT   CONFLICT   327   327       L -> I (IN REF. 2).
FT   CONFLICT   478   478       P -> G (IN REF. 2).
FT   CONFLICT   509   509       P -> D (IN REF. 2).
SQ   SEQUENCE   560 AA;  60959 MW;  FD65EDD2 CRC32;

  Query Match          26.6%;  Score 258;  DB 1;  Length 560;
  Best Local Similarity  32.8%;  Pred. No. 3.43e-22;
  Matches   43;  Conservative   37;  Mismatches  51;  Indels   0;  Gaps   0;

Db     76 LQRLMISDSHISAVAPGTFSDLIKLKTLRLSRNKITHLPGALLDKMVLLEQLFLDHNALR 135
          |: | : :::||::  |:| ||  |  |||:||::  :|  |:     | :| |  | ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    136 GIDQNMFQKLVNLQELALNQNQLDFLPASLFTNLENLKLLDLSGNNLTHLPKGLLGAQAK 195
          :| :: |:  |:: :| |: ||:  :  : |  | :| :| |: ||:|:|: : :   :|
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    196 LERLLLHSNRL 206
          |  : |||| |
Qy    122 LRTFRLHSNNL 132


RESULT    8
ID   ALS_HUMAN      STANDARD;      PRT;   605 AA.
AC   P35858;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 92357025.
RA   LEONG S.R., BAXTER R.C., CAMERATO T., DAI J., WOOD W.I.;
RT   "Structure and functional expression of the acid-labile subunit of
RT   the insulin-like growth factor-binding protein complex.";
RL   MOL. ENDOCRINOL. 6:870-876(1992).
RN   [2]
RP   SEQUENCE OF 28-35.
RX   MEDLINE; 89308584.
RA   BAXTER R.C., MARTIN J.L., BENIAC V.A.;
RT   "High molecular weight insulin-like growth factor binding protein
RT   complex. Purification and properties of the acid-labile subunit from
RT   human serum.";
RL   J. BIOL. CHEM. 264:11843-11848(1989).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
```

```
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86826; G184808; -.
DR   PIR; A41915; A41915.
DR   MIM; 601489; -.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL       1    27
FT   CHAIN       28   605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                              PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN      79   536       LEUCINE-RICH REPEATS.
FT   REPEAT      79    89       LRR 1.
FT   REPEAT      90   113       LRR 2.
FT   REPEAT     114   137       LRR 3.
FT   REPEAT     138   161       LRR 4.
FT   REPEAT     162   185       LRR 5.
FT   REPEAT     186   209       LRR 6.
FT   REPEAT     210   233       LRR 7.
FT   REPEAT     234   257       LRR 8.
FT   REPEAT     258   281       LRR 9.
FT   REPEAT     282   305       LRR 10.
FT   REPEAT     306   329       LRR 11.
FT   REPEAT     330   353       LRR 12.
FT   REPEAT     354   377       LRR 13.
FT   REPEAT     378   401       LRR 14.
FT   REPEAT     402   425       LRR 15.
FT   REPEAT     426   449       LRR 16.
FT   REPEAT     450   473       LRR 17.
FT   REPEAT     474   497       LRR 18.
FT   REPEAT     498   521       LRR 19.
FT   REPEAT     522   536       LRR 20.
FT   CARBOHYD    64    64       POTENTIAL.
FT   CARBOHYD    85    85       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   368   368       POTENTIAL.
FT   CARBOHYD   515   515       POTENTIAL.
FT   CARBOHYD   580   580       POTENTIAL.
SQ   SEQUENCE   605 AA;  66034 MW;  B5027E19 CRC32;

  Query Match          26.0%;  Score 252;  DB 1;  Length 605;
  Best Local Similarity  32.8%;  Pred. No. 2.91e-21;
  Matches   43;  Conservative   30;  Mismatches  58;  Indels   0;  Gaps   0;

Db    220 LRELDLSRNALRAIKANVFVQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279
          || |:|  | : :|    | :|  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHNRIRQLAERSFEGLGQ 339
          ::   :| | : :: |:|  | |: :   :|: |: || | | :| | :|:  ||: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLDHNQL 350
          | : |  | |
Qy    122 LRTFRLHSNNL 132


RESULT    9
ID   ALS_PAPPA      STANDARD;      PRT;   605 AA.
AC   O02833;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   PAPIO PAPIO (GUINEA BABOON).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97040714.
RA   DELHANTY P., BAXTER R.C.;
RT   "The cloning and expression of the baboon acid-labile subunit of the
RT   insulin-like growth factor binding protein complex.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 227:897-902(1996).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83462; E323796; -.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P23945; 1XUN.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL       1    27       BY SIMILARITY.
FT   CHAIN       28   605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                              PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN      79   536       LEUCINE-RICH REPEATS.
FT   REPEAT      79    89       LRR 1.
FT   REPEAT      90   113       LRR 2.
FT   REPEAT     114   137       LRR 3.
FT   REPEAT     138   161       LRR 4.
FT   REPEAT     162   185       LRR 5.
FT   REPEAT     186   209       LRR 6.
FT   REPEAT     210   233       LRR 7.
FT   REPEAT     234   257       LRR 8.
FT   REPEAT     258   281       LRR 9.
FT   REPEAT     282   305       LRR 10.
FT   REPEAT     306   329       LRR 11.
FT   REPEAT     330   353       LRR 12.
FT   REPEAT     354   377       LRR 13.
FT   REPEAT     378   401       LRR 14.
FT   REPEAT     402   425       LRR 15.
FT   REPEAT     426   449       LRR 16.
FT   REPEAT     450   473       LRR 17.
FT   REPEAT     474   497       LRR 18.
FT   REPEAT     498   521       LRR 19.
FT   REPEAT     522   536       LRR 20.
FT   CARBOHYD    64    64       POTENTIAL.
FT   CARBOHYD    85    85       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   368   368       POTENTIAL.
FT   CARBOHYD   515   515       POTENTIAL.
FT   CARBOHYD   580   580       POTENTIAL.
SQ   SEQUENCE   605 AA;  66110 MW;  5DF04D42 CRC32;

  Query Match          25.5%;  Score 248;  DB 1;  Length 605;
  Best Local Similarity  32.8%;  Pred. No. 1.20e-20;
  Matches   43;  Conservative   29;  Mismatches  59;  Indels   0;  Gaps   0;
```

```
Db    220 LRELDLSRNALRAIKANVFAQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279
          || |:|  | : :|    | :|  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQ 339
          ::   :| | : :: |:|  | |:: :  :|  |: || | | :| | :|:  ||: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLDHNQL 350
          |  : |  | |
Qy    122 LRTFRLHSNNL 132


RESULT  10
ID  ALS_RAT       STANDARD;      PRT;   603 AA.
AC  P35859;
DT  01-JUN-1994 (REL. 29, CREATED)
DT  01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE  PRECURSOR (ALS).
GN  IGFALS OR ALS.
OS  RATTUS NORVEGICUS (RAT).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=LIVER;
RX  MEDLINE; 93038676.
RA  DAI J., BAXTER R.C.;
RT  "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT  growth factor binding protein complex.";
RL  BIOCHEM. BIOPHYS. RES. COMMUN. 188:304-309(1992).
RN  [2]
RP  SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC  STRAIN=WISTAR; TISSUE=SERUM;
RX  MEDLINE; 94130835.
RA  BAXTER R.C., DAI J.;
RT  "Purification and characterization of the acid-labile subunit of rat
RT  serum insulin-like growth factor binding protein complex.";
RL  ENDOCRINOLOGY 134:848-852(1994).
CC  -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC      CIRCULATING IGFS TO THE TISSUES.
CC  -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC      OR IGF-II AND IGFBP-3.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC  -!- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC      AND LIVER.
CC  -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC      MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC  ----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ----------------------------------------------------------------------
DR  EMBL; S46785; E64972; -.
DR  PIR; JC1282; JC1282.
DR  PFAM; PF00560; LRR; 10.
DR  HSSP; P23945; 1XUN.
KW  GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT  SIGNAL       1     23
FT  CHAIN       24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                              PROTEIN, ACID LABILE CHAIN.
FT  DOMAIN      79    535       LEUCINE-RICH REPEATS.
FT  REPEAT      79     89       LRR 1.
FT  REPEAT      90    113       LRR 2.
FT  REPEAT     114    137       LRR 3.
FT  REPEAT     138    161       LRR 4.
FT  REPEAT     162    185       LRR 5.
FT  REPEAT     186    209       LRR 6.
FT  REPEAT     210    233       LRR 7.
FT  REPEAT     234    257       LRR 8.
FT  REPEAT     258    281       LRR 9.
FT  REPEAT     282    305       LRR 10.
FT  REPEAT     306    329       LRR 11.
FT  REPEAT     330    353       LRR 12.
FT  REPEAT     354    377       LRR 13.
FT  REPEAT     378    401       LRR 14.
FT  REPEAT     402    425       LRR 15.
FT  REPEAT     426    449       LRR 16.
FT  REPEAT     450    473       LRR 17.
FT  REPEAT     474    497       LRR 18.
FT  REPEAT     498    521       LRR 19.
FT  REPEAT     522    535       LRR 20.
FT  CARBOHYD    64     64       POTENTIAL.
FT  CARBOHYD    85     85       POTENTIAL.
FT  CARBOHYD    96     96       POTENTIAL.
FT  CARBOHYD   368    368       POTENTIAL.
FT  CARBOHYD   515    515       POTENTIAL.
FT  CARBOHYD   578    578       POTENTIAL.
FT  CARBOHYD   586    586       POTENTIAL.
SQ  SEQUENCE   603 AA;  66811 MW;  5BB22D53 CRC32;

  Query Match          24.9%;  Score 242;  DB 1;  Length 603;
  Best Local Similarity  34.1%;  Pred. No. 1.00e-19;
  Matches   46;  Conservative   34;  Mismatches  52;  Indels   3;  Gaps   3;

Db    240 HLPRLQKLYLDRNLITAVAPGAFLGMKALRWLDLSHNRVAGLMEDTFPGLLGLHVLRLAH 299
          ||  || | :: | |:::  |||  :| | :| |::| :  : |  | |   |  | |:
Qy      1 HLRVLQ-L-ME-NRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSE 57

Db    300 NAIASLRPRTFKDLHFLEELQLGHNRIRQLGERTFEGLGQLEVLTLNDNQITEVRVGAFS 359
          | | ::  ::|:    : :|||  |:|  : : :| :| :|||||||:| || : |::|:
Qy     58 NQIQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFN 117

Db    360 GLFNVAVMNLSGNCL 374
          : ::   | :| |
Qy    118 HMPKLRTFRLHSNNL 132


RESULT  11
ID  ALS_MOUSE     STANDARD;      PRT;   603 AA.
AC  P70389;
DT  01-NOV-1997 (REL. 35, CREATED)
DT  01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE  PRECURSOR (ALS).
GN  IGFALS OR ALS OR ALBS.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=129/SV;
RX  MEDLINE; 96413591.
RA  BOISCLAIR Y.R., SETO D., HSIEH S., HURST K.R., OOI G.T.;
RT  "Organization and chromosomal localization of the gene encoding the
RT  mouse acid labile subunit of the insulin-like growth factor binding
RT  complex.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 93:10028-10033(1996).
CC  -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC      CIRCULATING IGFS TO THE TISSUES.
CC  -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC      OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC  -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC      MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC  ----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66900; G1621613; -.
DR   MGD; MGI:107973; IGFALS.
DR   PFAM; PF00560; LRR; 10.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; SIGNAL.
FT   SIGNAL       1    23       BY SIMILARITY.
FT   CHAIN       24   603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                              PROTEIN, ACID LABILE CHAIN.
FT   DOMAIN      79   535       LEUCINE-RICH REPEATS.
FT   REPEAT      79    89       LRR 1.
FT   REPEAT      90   113       LRR 2.
FT   REPEAT     114   137       LRR 3.
FT   REPEAT     138   161       LRR 4.
FT   REPEAT     162   185       LRR 5.
FT   REPEAT     186   209       LRR 6.
FT   REPEAT     210   233       LRR 7.
FT   REPEAT     234   257       LRR 8.
FT   REPEAT     258   281       LRR 9.
FT   REPEAT     282   305       LRR 10.
FT   REPEAT     306   329       LRR 11.
FT   REPEAT     330   353       LRR 12.
FT   REPEAT     354   377       LRR 13.
FT   REPEAT     378   401       LRR 14.
FT   REPEAT     402   425       LRR 15.
FT   REPEAT     426   449       LRR 16.
FT   REPEAT     450   473       LRR 17.
FT   REPEAT     474   497       LRR 18.
FT   REPEAT     498   521       LRR 19.
FT   REPEAT     522   535       LRR 20.
FT   CARBOHYD    64    64       POTENTIAL.
FT   CARBOHYD    85    85       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   368   368       POTENTIAL.
FT   CARBOHYD   515   515       POTENTIAL.
FT   CARBOHYD   578   578       POTENTIAL.
FT   CARBOHYD   586   586       POTENTIAL.
SQ   SEQUENCE   603 AA;  66959 MW;  11ADB606 CRC32;

  Query Match          24.8%;  Score 241;  DB 1;  Length 603;
  Best Local Similarity  30.5%;  Pred. No. 1.42e-19;
  Matches   40;  Conservative   32;  Mismatches  59;  Indels   0;  Gaps   0;

      220 LRELDLSRNALRSVKANVFIHLPRLQKLYLDRNLITAVAPRAFLGMKALRWLDLSHNRVA 279
          || |:|  | : ::    |  |  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLLEDTFPGLLGLHVLRLAHNAITSLRPRTFKDLHFLEELQLGHNRIRQLGEKTFEGLGQ 339
          ::   :| | : :  |:|  | |: :   :|: |: || | | :| | :|:  :|: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLNDNQI 350
          |  : |: | :
Qy    122 LRTFRLHSNNL 132


RESULT  12
ID   GPCR_LYMST     STANDARD;      PRT;  1115 AA.
AC   P46023;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   G-PROTEIN COUPLED RECEPTOR GRL101 PRECURSOR.
OS   LYMNAEA STAGNALIS (GREAT POND SNAIL).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PULMONATA; BASOMMATOPHORA;
OC   LYMNAEIDAE; LYMNAEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CNS;
RX   MEDLINE; 94255418.
RA   TENSEN C.P., VAN KESTEREN E.R., PLANTA R.J., COX K.J.A., BURKE J.F.,
RA   VAN HEERIKHUIZEN H., VREUGDENHIL E.;
RT   "A G protein-coupled receptor with low density lipoprotein-binding
RT   motifs suggests a role for lipoproteins in G-linked signal
RT   transduction.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 91:4816-4820(1994).
CC   -!- FUNCTION: MIGHT DIRECTLY TRANSDUCE SIGNALS CARRIED BY LARGE
CC       EXTRACELLULAR (LIPO)PROTEIN(COMPLEXE)S INTO NEURONAL EVENTS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN A SMALL NUMBER OF
CC       NEURONS WITHIN THE CENTRAL NERVOUS SYSTEM AND TO A LESSER EXTENT
CC       IN THE HEART.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 12 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 6.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z23104; G438129; -.
DR   PIR; S40241; S40241.
DR   GCRDB; GCR_0860; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR   PROSITE; PS01209; LDLRA_1; 6.
DR   PROSITE; PS50068; LDLRA_2; 11.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00057; ldl_recept_a; 11.
DR   PFAM; PF00560; LRR; 3.
DR   HSSP; P01130; 1AJJ.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
KW   SIGNAL.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25  1115       G-PROTEIN COUPLED RECEPTOR GRL101.
FT   DOMAIN      25   767       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   768   788       1 (POTENTIAL).
FT   DOMAIN     789   801       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   802   822       2 (POTENTIAL).
FT   DOMAIN     823   857       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   858   878       3 (POTENTIAL).
FT   DOMAIN     879   887       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   888   908       4 (POTENTIAL).
FT   DOMAIN     909   941       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   942   962       5 (POTENTIAL).
FT   DOMAIN     963   988       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   989  1009       6 (POTENTIAL).
FT   DOMAIN    1010  1017       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1018  1038       7 (POTENTIAL).
FT   DOMAIN    1039  1115       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      32   523       12 X 40 AA APPROXIMATE TANDEM REPEATS
FT                              SIMILAR TO THE LDL-RECEPTOR CLASS A.
FT   DOMAIN      36    79       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      77   115       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     116   155       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     156   196       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     195   232       LDL-RECEPTOR CLASS A 5.
FT   DOMAIN     231   269       LDL-RECEPTOR CLASS A 6.
FT   DOMAIN     272   318       LDL-RECEPTOR CLASS A 7.
FT   DOMAIN     320   363       LDL-RECEPTOR CLASS A 8.
FT   DOMAIN     365   403       LDL-RECEPTOR CLASS A 9.
FT   DOMAIN     404   442       LDL-RECEPTOR CLASS A 10.
FT   DOMAIN     444   485       LDL-RECEPTOR CLASS A 11.
FT   DOMAIN     486   525       LDL-RECEPTOR CLASS A 12.
FT   DOMAIN     588   731       LEUCINE-RICH REPEATS.
FT   REPEAT     588   611       LRR 1.
FT   REPEAT     612   635       LRR 2.
FT   REPEAT     636   659       LRR 3.
FT   REPEAT     660   683       LRR 4.
```

```
FT   REPEAT      684    707       LRR 5.
FT   REPEAT      708    731       LRR 6.
FT   DISULFID     38     53       BY SIMILARITY.
FT   DISULFID     46     66       BY SIMILARITY.
FT   DISULFID     60     77       BY SIMILARITY.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     86    104       BY SIMILARITY.
FT   DISULFID     98    113       BY SIMILARITY.
FT   DISULFID    118    131       BY SIMILARITY.
FT   DISULFID    138    153       BY SIMILARITY.
FT   DISULFID    158    170       BY SIMILARITY.
FT   DISULFID    165    183       BY SIMILARITY.
FT   DISULFID    177    194       BY SIMILARITY.
FT   DISULFID    202    220       BY SIMILARITY.
FT   DISULFID    214    230       BY SIMILARITY.
FT   DISULFID    233    245       BY SIMILARITY.
FT   DISULFID    240    258       BY SIMILARITY.
FT   DISULFID    252    267       BY SIMILARITY.
FT   DISULFID    274    291       BY SIMILARITY.
FT   DISULFID    282    304       BY SIMILARITY.
FT   DISULFID    298    316       BY SIMILARITY.
FT   DISULFID    322    339       BY SIMILARITY.
FT   DISULFID    334    352       BY SIMILARITY.
FT   DISULFID    346    361       BY SIMILARITY.
FT   DISULFID    367    379       BY SIMILARITY.
FT   DISULFID    374    392       BY SIMILARITY.
FT   DISULFID    386    401       BY SIMILARITY.
FT   DISULFID    406    418       BY SIMILARITY.
FT   DISULFID    413    431       BY SIMILARITY.
FT   DISULFID    425    440       BY SIMILARITY.
FT   DISULFID    446    458       BY SIMILARITY.
FT   DISULFID    453    474       BY SIMILARITY.
FT   DISULFID    465    483       BY SIMILARITY.
FT   DISULFID    488    500       BY SIMILARITY.
FT   DISULFID    495    513       BY SIMILARITY.
FT   DISULFID    507    523       BY SIMILARITY.
FT   CARBOHYD     87     87       POTENTIAL.
FT   CARBOHYD    166    166       POTENTIAL.
FT   CARBOHYD    269    269       POTENTIAL.
FT   CARBOHYD    318    318       POTENTIAL.
FT   CARBOHYD    482    482       POTENTIAL.
FT   CARBOHYD    502    502       POTENTIAL.
FT   CARBOHYD    571    571       POTENTIAL.
FT   CARBOHYD    618    618       POTENTIAL.
FT   CARBOHYD    624    624       POTENTIAL.
FT   CARBOHYD    685    685       POTENTIAL.
SQ   SEQUENCE   1115 AA;  125865 MW;  2AEC245A CRC32;

  Query Match           24.4%;  Score 237;  DB 1;  Length 1115;
  Best Local Similarity  28.6%;  Pred. No. 5.79e-19;
  Matches   38;  Conservative   40;  Mismatches  55;  Indels   0;  Gaps   0;

Db    609 LTHLNLADNNITSLKNGSLLGLSNLKQLHINGNKIETIEEDTFSSMIHLTVLDLSNQRLT 668
          |  | | :| |:::  |::  | :| :|::| |::: : |  | :  :|  ||||: ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    669 HVYKNMFKGLKQITVLNISRNQINSIDNGAFNNLANVRLIDLSGNVIKDIGQKVFMGLPR 728
           : :: |:|  :|  | :  |||: |::||| | :: :: |: | |  ::   |  :|:
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    729 LVELKTDSYRFCC 741
          |  ::  |  : |
Qy    122 LRTFRLHSNNLYC 134


RESULT   13
ID   A2GL_HUMAN     STANDARD;      PRT;   312 AA.
AC   P02750;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85166241.
RA   TAKAHASHI N., TAKAHASHI Y., PUTNAM F.W.;
RT   "Periodicity of leucine and tandem repetition of a 24-amino acid
RT   segment in the primary structure of leucine-rich alpha 2-glycoprotein
RT   of human serum.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:1906-1910(1985).
CC   -!- FUNCTION: THE FUNCTION OF THIS PLASMA PROTEIN IS NOT KNOWN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS.
DR   PIR; A03211; NBHUA2.
DR   SWISS-2DPAGE; P02750; HUMAN.
DR   PFAM; PF00560; LRR; 4.
KW   PLASMA; GLYCOPROTEIN; REPEAT; LEUCINE-REPEAT.
FT   DISULFID      8     21
FT   DISULFID    268    294
FT   CARBOHYD      2      2
FT   CARBOHYD     44     44
FT   CARBOHYD    151    151
FT   CARBOHYD    234    234
FT   CARBOHYD    290    290
FT   CARBOHYD    271    271       POTENTIAL.
SQ   SEQUENCE   312 AA;  34346 MW;  48C3DB08 CRC32;

  Query Match           24.2%;  Score 235;  DB 1;  Length 312;
  Best Local Similarity  31.8%;  Pred. No. 1.17e-18;
  Matches   42;  Conservative   39;  Mismatches  49;  Indels   2;  Gaps   2;

Db    107 LDTLVLKENQLEVLEVSWLHGLKALGHLDLSGNRLRKLPPGL-LANFTLLRTLDLGENQL 165
          | | |:||::  :| : :: || | :| |: | |: :|  | |:   | | |||:|||:
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYR-LDLSENQI 60

Db    166 ETLPPDLLRGPLQLERLHLEGNKLQVLGKDLLLPQPDLRYLFLNGNKLARVAAGAFQGLR 225
          :::|   :||::::  |:|: | :  :    : :  ||  | || |:::|:: ::|  :
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    226 QLDMLDLSNNSL 237
          |  : | :|:|
Qy    121 KLRTFRLHSNNL 132


RESULT   14
ID   CHAD_BOVIN     STANDARD;      PRT;   361 AA.
AC   Q27972;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CHONDROADHERIN PRECURSOR (CARTILAGE LEUCINE-RICH PROTEIN) (38 KD BONE
DE   PROTEIN).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=CARTILAGE;
RX   MEDLINE; 94342341.
RA   NEAME P.J., SOMMARIN Y., BOYNTON R.E., HEINEGARD D.;
RT   "The structure of a 38-kDa leucine-rich protein (chondroadherin)
RT   isolated from bovine cartilage.";
RL   J. BIOL. CHEM. 269:21547-21554(1994).
RN   [2]
RP   SEQUENCE OF 25-55 AND 77-97.
RC   TISSUE=BONE;
RX   MEDLINE; 95113864.
RA   HU B., COULSON L., MOYER B., PRICE P.A.;
RT   "Isolation and molecular cloning of a novel bone phosphoprotein
RT   related in sequence to the cystatin family of thiol protease
RT   inhibitors.";
RL   J. BIOL. CHEM. 270:431-436(1995).
CC   ----------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08018; G470672; -.
DR   PFAM; PF00560; LRR; 5.
KW   REPEAT; SIGNAL.
FT   SIGNAL       1    24       OR 23 (IN SOME ISOFORM(S)).
FT   CHAIN       25   361       CHONDROADHERIN.
FT   CHAIN       25   352       CHONDROADHERIN, MINOR FORM.
FT   DOMAIN      79   317       10 X 24 AA LEUCINE-RICH TANDEM REPEATS.
FT   REPEAT      79   102       1.
FT   REPEAT     103   126       2.
FT   REPEAT     127   150       3.
FT   REPEAT     151   174       4.
FT   REPEAT     175   198       5.
FT   REPEAT     199   222       6.
FT   REPEAT     223   246       7.
FT   REPEAT     248   271       8.
FT   REPEAT     272   293       9.
FT   REPEAT     294   317       10.
FT   DISULFID   306   348
FT   DISULFID   308   328
FT   CONFLICT    25    25       C -> Y (IN REF. 2).
FT   CONFLICT    29    29       C -> W (IN REF. 2).
FT   CONFLICT    31    31       C -> H (IN REF. 2).
FT   CONFLICT    40    40       C -> L (IN REF. 2).
FT   CONFLICT    52    52       S -> R (IN REF. 2).
SQ   SEQUENCE   361 AA;  40884 MW;  A370BB91 CRC32;

  Query Match          22.9%;  Score 222;  DB 1;  Length 361;
  Best Local Similarity  33.6%;  Pred. No. 1.07e-16;
  Matches   45;  Conservative   30;  Mismatches  56;  Indels   3;  Gaps   3;

Db    150 NLFILQLNNNKIRELRSGAFQGAKDLRWLYLSENSLSSLQPGAL-DDVENLAKFYLDRNQ 208
          :| :||| :|:|  :  |||| |:| :| |: |:|  | |  |      | :: |  ||
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQ-LFPELLFLGTARLYRLDLSENQ 59

Db    209 LSSYPSAALSKLRVVEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFSDGAFLG 268
          : : |  |:     : :| |  | :  | | ||::: | || | |:| |: ::| ::|
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRAL-RDLEVLTLNNNNITRLSVASFNH 118

Db    269 VTTLKHVHLENNRL 282
          :  |:  :| :| |
Qy    119 MPKLRTFRLHSNNL 132


RESULT   15
ID   GARP_HUMAN     STANDARD;      PRT;   662 AA.
AC   Q14392;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GARP PROTEIN PRECURSOR (GARPIN).
GN   GARP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94235567.
RA   OLLENDORFF V., NOGUCHI T., DELAPEYRIERE O., BIRNBAUM D.;
RT   "The GARP gene encodes a new member of the family of leucine-rich
RT   repeat-containing proteins.";
RL   CELL GROWTH DIFFER. 5:213-219(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20, THERE ARE TWO BLOCKS OF
CC       10 LRR'S.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z24680; G439296; -.
DR   MIM; 137207; -.
DR   PFAM; PF00560; LRR; 10.
KW   GLYCOPROTEIN; LEUCINE-REPEAT; REPEAT; TRANSMEMBRANE; SIGNAL.
FT   SIGNAL       1    19       POTENTIAL.
FT   CHAIN       20   662       GARP PROTEIN.
FT   DOMAIN      20   627       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   628   648       POTENTIAL.
FT   DOMAIN     649   662       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      54   292       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     320   565       LEUCINE-RICH REPEATS (2ND REGION).
FT   REPEAT      54    77       LRR 1-1.
FT   REPEAT      78   101       LRR 1-2.
FT   REPEAT     102   128       LRR 1-3.
FT   REPEAT     129   153       LRR 1-4.
FT   REPEAT     154   177       LRR 1-5.
FT   REPEAT     178   201       LRR 1-6.
FT   REPEAT     202   222       LRR 1-7.
FT   REPEAT     223   247       LRR 1-8.
FT   REPEAT     248   269       LRR 1-9.
FT   REPEAT     270   292       LRR 1-10.
FT   REPEAT     320   343       LRR 2-1.
FT   REPEAT     344   367       LRR 2-2.
FT   REPEAT     368   390       LRR 2-3.
FT   REPEAT     391   414       LRR 2-4.
FT   REPEAT     415   438       LRR 2-5.
FT   REPEAT     448   470       LRR 2-6.
FT   REPEAT     471   495       LRR 2-7.
FT   REPEAT     496   518       LRR 2-8.
FT   REPEAT     519   540       LRR 2-9.
FT   REPEAT     541   565       LRR 2-10.
FT   CARBOHYD   203   203       POTENTIAL.
FT   CARBOHYD   271   271       POTENTIAL.
FT   CARBOHYD   308   308       POTENTIAL.
FT   CARBOHYD   345   345       POTENTIAL.
FT   CARBOHYD   545   545       POTENTIAL.
SQ   SEQUENCE   662 AA;  71978 MW;  D7B74960 CRC32;

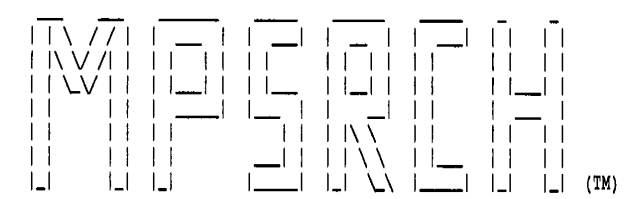
  Query Match          22.9%;  Score 222;  DB 1;  Length 662;
  Best Local Similarity  32.8%;  Pred. No. 1.07e-16;
  Matches   45;  Conservative   30;  Mismatches  58;  Indels   4;  Gaps   4;

Db     75 LRHLDLSTNEISFLQPGAFQALTHLEHLSLAHNRLAMATALSAGGLGPLPRVTSLDLSGN 134
          || |:|  | || :: |||| |  ||:| | :| | :  |    ||  :|:  |||| |
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPEL-LF-LGT-ARLYRLDLSEN 58

Db    135 SLYSGLLERLLGEAPSLHTLSLAENSLTRLTRHTFRDMPALEQLDLHSNVLMDIEDGAFE 194
          :  ::  : :  |  :  | |  | :: :   :|| :  || | |::| :  :  ::|:
Qy     59 QI-QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFN 117

Db    195 GLPRLTHLNLSRNSLTC 211
          :|:|  : |  |:| |
Qy    118 HMPKLRTFRLHSNNLYC 134


Search completed: Fri May 28 09:30:54 1999
Job time : 18 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:31:17 1999; MasPar time 14.13 Seconds
517.680 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-12
Description: (1-134) from US09191647.pep
Perfect Score: 971
Sequence: 1 HLRVLQLMENRISTIERGAF.....SFNHMPKLRTFRLHSNNLYC 134

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 44.205; Variance 111.499; scale 0.396

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	780	80.3	1523	11	O88280	MEGF5.	1.47e-102
2	777	80.0	1531	11	O88279	MEGF4.	4.52e-102
3	347	35.7	1496	4	Q92626	MYELOBLAST KIAA0230 (F	8.67e-34
4	274	28.2	1066	5	Q18902	CODED FOR BY C. ELEGAN	6.72e-23
5	270	27.8	811	4	O75139	KIAA0644 PROTEIN.	2.58e-22
6	267	27.5	610	5	Q21604	M88.6 PROTEIN.	7.07e-22
7	267	27.5	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	7.07e-22
8	261	26.9	907	4	O75473	ORPHAN G PROTEIN-COUPL	5.26e-21
9	260	26.8	331	13	O93233	PHOSPHOLIPASE A2 INHIB	7.34e-21
10	254	26.2	1385	5	Q26388	TLR=TOLL-LIKE RECEPTOR	5.41e-20
11	250	25.7	1389	5	Q24591	WHEELER.	2.04e-19
12	246	25.3	738	5	Q93373	C44H4.2 PROTEIN.	7.65e-19
13	245	25.2	733	5	Q24250	TARTAN PROTEIN PRECURS	1.06e-18
14	241	24.8	603	11	O70211	INSULIN-LIKE GROWTH FA	3.97e-18
15	239	24.6	683	5	Q22187	T05A1.3 PROTEIN.	7.66e-18
16	238	24.5	880	5	P91643	KEK1 PRECURSOR.	1.06e-17
17	235	24.2	358	11	O70210	CHONDROADHERIN.	2.84e-17
18	234	24.1	358	11	O55226	CHONDROADHERIN.	3.94e-17
19	232	23.9	562	5	Q21164	SIMILAR TO LRR.	7.57e-17
20	230	23.7	428	4	O14498	ISLR PRECURSOR.	1.45e-16
21	230	23.7	516	4	O43300	KIAA0416.	1.45e-16
22	230	23.7	660	4	O43155	KIAA0405.	1.45e-16
23	226	23.3	1100	5	Q24622	TOLL PROTEIN.	5.33e-16
24	224	23.1	789	5	O16781	SIMILARITY TO MULTIPLE	1.02e-15
25	224	23.1	1355	5	O16779	CODED FOR BY C. ELEGAN	1.02e-15
26	222	22.9	713	4	O75325	GLIOMA AMPLIFIED ON CH	1.94e-15
27	218	22.5	892	5	P91644	KEK2 PRECURSOR (FRAGME	7.06e-15
28	217	22.3	653	5	O02329	T23G11.6 PROTEIN.	9.73e-15
29	215	22.1	359	4	O15335	CHONDROADHERIN.	1.85e-14
30	213	21.9	716	11	Q61809	LEUCINE-RICH-REPEAT PR	3.51e-14
31	213	21.9	718	13	O73675	NEURONAL LEUCINE-RICH	3.51e-14
32	211	21.7	135	6	O46377	BIGLYCAN (FRAGMENT).	6.65e-14
33	211	21.7	224	5	O44086	ZK994.4 PROTEIN.	6.65e-14
34	210	21.6	369	6	O46390	BIGLYCAN PRECURSOR.	9.15e-14
35	210	21.6	372	6	O46403	BIGLYCAN.	9.15e-14
36	209	21.5	707	11	P97860	LEUCINE-RICH REPEAT PR	1.26e-13
37	209	21.5	961	5	P90920	K07A12.2 PROTEIN.	1.26e-13
38	206	21.2	526	10	O22753	PREDICTED LEUCINE-RICH	3.27e-13
39	206	21.2	680	5	Q93374	C44H4.3 PROTEIN.	3.27e-13
40	203	20.9	458	5	Q93377	C44H4.1 PROTEIN.	8.46e-13
41	203	20.9	1535	5	Q23991	PEROXIDASIN PRECURSOR.	8.46e-13
42	201	20.7	360	6	O46542	DERMATAN SULFATE PROTE	1.59e-12
43	201	20.7	360	6	Q28888	DECORIN.	1.59e-12
44	201	20.7	522	4	O43354	BAC CLONE GS099H08, CO	1.59e-12
45	200	20.6	515	5	O15912	RANDOM SLUG CDNA21 PRO	2.18e-12

ALIGNMENTS

RESULT 1
ID O88280 PRELIMINARY; PRT; 1523 AA.
AC O88280;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF5.
GN MEGF5.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011531; D1033424; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1523 AA; 167767 MW; 2BD845D0 CRC32;

Query Match 80.3%; Score 780; DB 11; Length 1523;
Best Local Similarity 76.1%; Pred. No. 1.47e-102;
Matches 102; Conservative 23; Mismatches 9; Indels 0; Gaps 0;

```
Db    86 NLRVLHLEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKLTRLDLSENQI 145
         :||||:| :|::| ||||||||||:|||||||:|:||::||||| :|:::| |||||||||
Qy     1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db   146 QGIPRKAFRGVTGVKNLQLDNNHISCIEDGAFRALRDLEILTLNNNNISRILVTSFNHMP 205
         |:|||||||   :|||||| |:|||||||||||||||:||||||||:|: |:||||||
Qy    61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db   206 KIRTLRLHSNHLYC 219
         |:||:|||||:|||
Qy   121 KLRTFRLHSNNLYC 134
```

RESULT 2

```
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match           80.0%;  Score 777;  DB 11;  Length 1531;
  Best Local Similarity  77.6%;  Pred. No. 4.52e-102;
  Matches  104;  Conservative   18;  Mismatches  12;  Indels   0;  Gaps   0;

Db     86 QLRVLQLMENQIGAVERGAFDDMKELERLRLNRNQLQVLPELLFQNNQALSRLDLSENSL 145
           :|||||||||:|:::|||||:|:||||||||||| ||::|||||     | |||||| :
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    146 QAVPRKAFRGATDLKNLQLDKNQISCIEEGAFRALRGLEVLTLNNNNITTIPVSSFNHMP 205
           ||:|||||||| |:|||||| |||||||:||||||| |||||||||||| ::|:||||||
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    206 KLRTFRLHSNHLFC 219
           ||||||||||:|:|
Qy    121 KLRTFRLHSNNLYC 134


RESULT    3
ID   Q92626      PRELIMINARY;      PRT;  1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELOBLAST KIAA0230 (FRAGMENT).
GN   KIAA0230.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA RES. 3:321-329(1996).
DR   EMBL; D86983; D1013908; -.
DR   PFAM; PF00047; ig; 4.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00141; peroxidase; 1.
DR   PFAM; PF00560; LRR; 3.
FT   NON_TER       1      1
SQ   SEQUENCE   1496 AA;  167209 MW;  5731EE51 CRC32;

  Query Match           35.7%;  Score 347;  DB 4;  Length 1496;
  Best Local Similarity  40.9%;  Pred. No. 8.67e-34;
  Matches   54;  Conservative   26;  Mismatches  50;  Indels   2;  Gaps   2;

Db     83 ILDLRFNRIREIQPGAFRRLRNLNTLLLNNNQIKRIPSGAFEDLENLKY-LYLYKNEIQS 141
           :|:|  |||  |: |||: |::|: | || | :  :|   |     | | | |  |:||:
Qy      4 VLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARL-YRLDLSENQIQA 62

Db    142 IDRQAFKGLASLEQLYLHFNQIETLDPDSFQHLPKLERLFLHNNRITHLVPGTFNHLESM 201
           | | ||:|   :  | | :|||  ::  :|: |  || | |:|| ||:|  ::|||:  :
Qy     63 IPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKL 122

Db    202 KRLRLDSNTLHC 213
           : :|| || | |
Qy    123 RTFRLHSNNLYC 134


RESULT    4
ID   Q18902      PRELIMINARY;      PRT;  1066 AA.
AC   Q18902;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CODED FOR BY C. ELEGANS CDNA YK132E5.5.
GN   C56E6.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FULTON L.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U39996; G1055120; -.
DR   PFAM; PF00560; LRR; 18.
SQ   SEQUENCE   1066 AA;  122109 MW;  20D7DDEF CRC32;

  Query Match           28.2%;  Score 274;  DB 5;  Length 1066;
  Best Local Similarity  31.6%;  Pred. No. 6.72e-23;
  Matches   42;  Conservative   37;  Mismatches  52;  Indels   2;  Gaps   2;

Db    392 LRHLMLDNNQIQKIDNFSLADLPKLQHLSLAGNQLDIITENMFGSSSSSELKSLNLAHNK 451
           || | | :|:|  |:  :: ||  |::| |  | |::: | :|  :::  |  |:|: |
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLF-LGTAR-LYRLDLSENQ 59

Db    452 IHSISSRSFSDLDNLQQLRLSHNNIRTITSMTFSNLRNLRYLDLSHNRIIKILPSALYQL 511
           |::|: ::|   ::  |:|  | |  |   :|  ||:|  | |::| | ::  ::: ::
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    512 PALDVLHLDHNNL 524
           | |  ::|  |||
Qy    120 PKLRTFRLHSNNL 132


RESULT    5
ID   O75139      PRELIMINARY;      PRT;   811 AA.
```

```
AC   O75139;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KIAA0644 PROTEIN.
GN   KIAA0644.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98403880.
RA   ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA   NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA RES. 5:169-176(1998).
DR   EMBL; AB014544; D1032580; -.
SQ   SEQUENCE   811 AA;  88695 MW;  C8B8C147 CRC32;

  Query Match           27.8%;  Score 270;  DB 4;  Length 811;
  Best Local Similarity  32.3%;  Pred. No. 2.58e-22;
  Matches   43;  Conservative   34;  Mismatches  54;  Indels   2;  Gaps   2;

Db    133 LRILYANGNEISRLSRGSFEGLESLVKLRLDGNALGALPDAVFAPLGNLLYLHLESNRIR 192
          ||:|    | || : ||:|: |  | :|||: | |  :|: :|   : |  | |  |:|:
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    193 FLGKNAFAQLGKLRFLNLSANELQPSLRHAATFAPLRSLSSLILSANSLQHLGPRIFQHL 252
           : ::||     :: | |  |:: : :   ::| :|| |  | |: |:: :|:   | |:
Qy     62 AIPRKAFRGAVDIKNLQLDYNQI-SCI-EDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    253 PRLGLLSLRGNQL 265
          |:|  : |::| |
Qy    120 PKLRTFRLHSNNL 132


RESULT   6
ID   Q21604      PRELIMINARY;      PRT;   610 AA.
AC   Q21604;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   M88.6 PROTEIN.
GN   M88.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SULSTON J.;
RL   SUBMITTED (JUN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z34802; E1348350; -.
SQ   SEQUENCE   610 AA;  68394 MW;  515CAA24 CRC32;

  Query Match           27.5%;  Score 267;  DB 5;  Length 610;
  Best Local Similarity  35.5%;  Pred. No. 7.07e-22;
  Matches   43;  Conservative   28;  Mismatches  50;  Indels   0;  Gaps   0;

Db    153 LKTLDLAMNKIQEIDVGAFEELKKVEELLLNENDIRVLKTGTFDGMKNLKKLTLQNCNLE 212
          |: |:|  |:|  |: |||::|| :| | || |:::::    | |   | :| | :  ::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    213 IIQKGAFRGLNSLEQLILSNNNLENIDWTIFSALKNLRVLDLGSNKISNVEMKSFPKLEK 272
           | : ||||   :  | |  | :  |:   | ||::| || | :|:|: : : ||  : |
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    273 L 273
          |
Qy    122 L 122


RESULT   7
ID   P70193      PRELIMINARY;      PRT;  1091 AA.
AC   P70193;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96394313.
RA   SUZUKI Y., SATO N., TOHYAMA M., WANAKA A., TAKAGI T.;
RT   "cDNA cloning of a novel membrane glycoprotein that is expressed
RT   specifically in glial cells in the mouse brain LIG-1: A protein with
RT   leucine-rich repeats and immunoglobulin-like domains.";
RL   J. BIOL. CHEM. 271:22522-22527(1996).
DR   EMBL; D78572; D1012081; -.
DR   MGD; MGI:107935; IMG.
DR   PFAM; PF00047; ig; 3.
DR   PFAM; PF00560; LRR; 7.
KW   MEMBRANE.
SQ   SEQUENCE   1091 AA;  119283 MW;  C0F262F9 CRC32;

  Query Match           27.5%;  Score 267;  DB 11;  Length 1091;
  Best Local Similarity  35.3%;  Pred. No. 7.07e-22;
  Matches   47;  Conservative   37;  Mismatches  47;  Indels   2;  Gaps   2;

Db    166 RIRELNLASNRISILESGAFDGLSRSLLTLRLSKNRITQLPVKAF-KLPRLTQLDLNRNR 224
          ::| | |  |||| :| |||: | : |  |||::| :  :|   |   :|| :|||: |:
Qy      1 HLRVLQLMENRISTIERGAFQDL-KELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQ 59

Db    225 IRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVNSGSLYGL 284
          |: |   :|:|   :  |:|: | || : ||||::|  : || |: |::  :: :|:  :
Qy     60 IQAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHM 119

Db    285 TALHQLHLSNNSI 297
            |: ::| :|::
Qy    120 PKLRTFRLHSNNL 132


RESULT   8
ID   O75473      PRELIMINARY;      PRT;   907 AA.
AC   O75473;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ORPHAN G PROTEIN-COUPLED RECEPTOR HG38.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98308104.
RA   MCDONALD T., WANG R., BAILEY W., XIE G., CHEN F., CASKEY C.T.,
```

```
RA   LIU Q.;
RT   "Identification and cloning of an orphan G protein-coupled receptor
RT   of the glycoprotein hormone receptor subfamily.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 247:266-270(1998).
DR   EMBL; AF062006; G3366802; -.
SQ   SEQUENCE   907 AA;  99997 MW;  B9147406 CRC32;

  Query Match           26.9%;  Score 261;  DB 4;  Length 907;
  Best Local Similarity  36.6%;  Pred. No. 5.26e-21;
  Matches   48;  Conservative   24;  Mismatches  59;  Indels   0;  Gaps   0;

Db    116 LKVLMLQNNQLRHVPTEALQNLRSLQSLRLDANHISYVPPSCFSGLHSLRHLWLDDNALT 175
          |:|| | :|::  :   |:|:|: |: |||: |::   |   | |   | :| | :| :
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    176 EIPVQAFRSLSALQAMTLALNKIHHIPDYAFGNLSSLVVLHLHNNRIHSLGKKCFDGLHS 235
           ||  |||:   :  : |  | |  | | ||  |  | || |:|| |  |:   |: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

       236 LETLDLNYNNL 246
           | |: |: |||
Qy     122 LRTFRLHSNNL 132


RESULT   9
ID   O93233      PRELIMINARY;      PRT;   331 AA.
AC   O93233;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PHOSPHOLIPASE A2 INHIBITOR.
OS   GLOYDIUS BLOMHOFFII (MAMUSHI).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; LEPIDOSAURIA; SQUAMATA;
OC   SCLEROGLOSSA; SERPENTES; COLUBROIDEA; VIPERIDAE; CROTALINAE;
OC   AGKISTRODON.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 98344034.
RA   OKUMURA K., OHKURA N., INOUE S., IKEDA K., HAYASHI K.;
RT   "A novel phospholipase A2 inhibitor with leucine-rich repeats from
RT   the blood plasma of Agkistrodon blomhoffii siniticus (subtitle:
RT   Sequence homologies with human leucine-rich alpha2-glycoprotein).";
RL   J. BIOL. CHEM. 273:19469-19475(1998).
DR   EMBL; AB007198; D1032956; -.
KW   PHOSPHOLIPASE A2 INHIBITOR.
SQ   SEQUENCE   331 AA;  37091 MW;  D764D70F CRC32;

  Query Match           26.8%;  Score 260;  DB 13;  Length 331;
  Best Local Similarity  32.8%;  Pred. No. 7.34e-21;
  Matches   44;  Conservative   33;  Mismatches  57;  Indels   0;  Gaps   0;

Db     80 NLQELHLSNNRLKTLPSGLFRNLPQLHTLDLSTNHLEDLPPEIFTNASSLILLPLSENQL 139
          :|: |:| :||: |:  | |::| :|  | |: |:|: :|  :|  :: |  | |||||:
Qy      1 HLRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQI 60

Db    140 AELHPSWFQTLGELRILGLDHNQVKEIPISCFDKLKKLTSLDLSFNLLRRLAPEMFSGLD 199
           :   |:   ::: | || ||:  |  : :  |: |  | |: | : ||:   |: :
Qy     61 QAIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMP 120

Db    200 NLEKLILESNPIQC 213
          :|  : | || : |
Qy    121 KLRTFRLHSNNLYC 134


RESULT   10
ID   Q26388      PRELIMINARY;      PRT;  1385 AA.
AC   Q26388;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TLR=TOLL-LIKE RECEPTOR.
GN   TLR.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95151581.
RA   CHIANG C., BEACHY P.A.;
RT   "Expression of a novel Toll-like gene spans the parasegment boundary
RT   and contributes to hedgehog function in the adult eye of
RT   Drosophila.";
RL   MECH. DEV. 47:225-239(1994).
DR   EMBL; S76155; G913248; -.
DR   FLYBASE; FBgn0004364; 18w.
DR   PFAM; PF00560; LRR; 13.
SQ   SEQUENCE   1385 AA;  154848 MW;  60273533 CRC32;

  Query Match           26.2%;  Score 254;  DB 5;  Length 1385;
  Best Local Similarity  32.1%;  Pred. No. 5.41e-20;
  Matches   42;  Conservative   33;  Mismatches  55;  Indels   1;  Gaps   1;

Db    359 LQILDMRNNSIGHIEEGAFLPLYNLHTLNLAENRLHTLDNRIFNGLYVLTKLTLNNNLVS 418
          |::|:: :| |: || |||  | :|  | |  | |: : : :| |   | :| |::| :
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    419 IVESQAFRNCSDLKELDLSSNQLTEVPEAA-QDLSMLKTLDLGENQISEFKNNTFRNLNQ 477
           :   |||   |:|:|:|  ||:: : ::| : |  |  | | :| |: :   :| ::
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    478 LTGLRLIDNRI 488
          |  :||  | :
Qy    122 LRTFRLHSNNL 132


RESULT   11
ID   Q24591      PRELIMINARY;      PRT;  1389 AA.
AC   Q24591;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   WHEELER.
GN   WHEELER.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95324375.
RA   ELDON E., KOOYER S., D'EVELYN D., DUMAN M., LAWINGER P., BOTAS J.,
RA   BELLEN H.;
RT   "The Drosophila 18 wheeler is required for morphogenesis and has
RT   striking similarities to Toll.";
RL   DEVELOPMENT 120:885-899(1994).
DR   EMBL; L23171; G1019104; -.
DR   FLYBASE; FBgn0004364; 18w.
DR   PFAM; PF00560; LRR; 13.
SQ   SEQUENCE   1389 AA;  155260 MW;  8F0E6F5A CRC32;

  Query Match           25.7%;  Score 250;  DB 5;  Length 1389;
  Best Local Similarity  31.3%;  Pred. No. 2.04e-19;
  Matches   41;  Conservative   33;  Mismatches  56;  Indels   1;  Gaps   1;

Db    359 LQILDMRNNSIGHIEEGAFLPLYNLHTLNLAENRLHTLDNRIFNGLYVLTKLTLNNNLVS 418
          |::|:: :| |: || |||  | :|  | |  | |: : : :| |   | :| |::| :
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    419 IVESQAFRNCSDLKELDLSSNQLTEVPEAV-QDLSMLKTLDLGENQISEFKNNTFRNLNQ 477
           :   |||   |:|:|:|  ||:: : ::  : |  |  | | :| |: :   :| ::
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    478 LTGLRLIDNRI 488
          |  :||  | :
```

```
Qy    122 LRTFRLHSNNL 132


RESULT  12
ID   Q93373      PRELIMINARY;      PRT;   738 AA.
AC   Q93373;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   C44H4.2 PROTEIN.
GN   C44H4.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMYE R.;
RL   SUBMITTED (AUG-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z79598; E1344562; -.
SQ   SEQUENCE   738 AA;  81830 MW;  868D9B70 CRC32;

  Query Match           25.3%;  Score 246;  DB 5;  Length 738;
  Best Local Similarity  29.0%;  Pred. No. 7.65e-19;
  Matches    38;  Conservative   41;  Mismatches  52;  Indels   0;  Gaps   0;

Db    151 IQTINLGHNNMTAVPSSAIRGLKQLQSLHLHKNRIEQLDALNFLNLPVLNLLNLAGNQIH 210
          ::  :  |  | ::::  :|:: ||:|: |:|::| :: :  | ||  : |  |:|: |||:
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    211 ELNRQAFLNVPSLRYLYLSGNKITKLTAYQFQTFEQLEMLDLTNNEIGAIPANSLSGLKQ 270
           :  | ||     :: | |  | |: :    |::: :||:| | ||:|  ::  |:: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    271 LRQLYLAHNKI 281
          || : |  |::
Qy    122 LRTFRLHSNNL 132


RESULT  13
ID   Q24250      PRELIMINARY;      PRT;   733 AA.
AC   Q24250;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TARTAN PROTEIN PRECURSOR.
GN   TRN.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 94074761.
RA   CHANG Z., PRICE B.D., BOCKHEIM S., BOEDIGHEIMER M.J., SMITH R.,
RA   LAUGHON A.;
RT   "Molecular and genetic characterization of the Drosophila tartan
RT   gene.";
RL   DEV. BIOL. 160:315-332(1993).
DR   EMBL; U02078; G408375; -.
DR   FLYBASE; FBgn0010452; trn.
DR   PFAM; PF00560; LRR; 6.
KW   SIGNAL.
FT   SIGNAL       1     24       POTENTIAL.
SQ   SEQUENCE   733 AA;  81319 MW;  DA426BB0 CRC32;

  Query Match           25.2%;  Score 245;  DB 5;  Length 733;
  Best Local Similarity  34.8%;  Pred. No. 1.06e-18;
  Matches    46;  Conservative   31;  Mismatches  54;  Indels   1;  Gaps   1;

Db    201 LAELFLGMNTLQSIQAGAFQDLKGLTRLELKGASLRNISHDSFLGLQELRILDLSDNRLD 260
          |  | |  | : :|: ||||||| | || |:  :|: ::   |||   |  ||||:|:::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    261 RIPSVGLSKLVRLEQLSLGQNDFEVISEGAFMGLKQLKRLEVNGALRLKRVMTGAFSDNG 320
           ||  ::   | :  | |  |::  | :||| :|::|  | :|    : |:  ::|:
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNN-NITRLSVASFNHMP 120

Db    321 NLEYLNLSSNKM 332
          :|  : | ||::
Qy    121 KLRTFRLHSNNL 132


RESULT  14
ID   O70211      PRELIMINARY;      PRT;   603 AA.
AC   O70211;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID-LABILE
DE   SUBUNIT.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER;
RX   MEDLINE; 98121980.
RA   DELHANTY P.J., BAXTER R.C.;
RT   "Cloning and characterization of the rat gene for the acid-labile
RT   subunit of the insulin-like growth factor binding protein complex.";
RL   J. MOL. ENDOCRINOL. 19:267-277(1997).
DR   EMBL; AF006203; G3093474; -.
SQ   SEQUENCE   603 AA;  66924 MW;  74E63165 CRC32;

  Query Match           24.8%;  Score 241;  DB 11;  Length 603;
  Best Local Similarity  30.5%;  Pred. No. 3.97e-18;
  Matches    40;  Conservative   32;  Mismatches  59;  Indels   0;  Gaps   0;

Db    220 LRELDLSRNALRSVKANVFVHLPRLQKLYLDRNLITAVAPRAFLGMKALRWLDLSHNRVA 279
          || |:|  | : ::    |  |  |::| |:|| :   :   |||   | :|||| |::
Qy      2 LRVLQLMENRISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQ 61

Db    280 GLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGERTFEGLGQ 339
          ::   :| | : :  |:|  | |: :   :|: |: || | | :| | :|:  :|: :
Qy     62 AIPRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPK 121

Db    340 LEVLTLNDNQI 350
          |  : |: | :
Qy    122 LRTFRLHSNNL 132


RESULT  15
ID   Q22187      PRELIMINARY;      PRT;   683 AA.
AC   Q22187;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   T05A1.3 PROTEIN.
GN   T05A1.3.
```

OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA LLOYD C.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z68219; E1349116; -.
SQ SEQUENCE 683 AA; 77437 MW; B9B00EA2 CRC32;

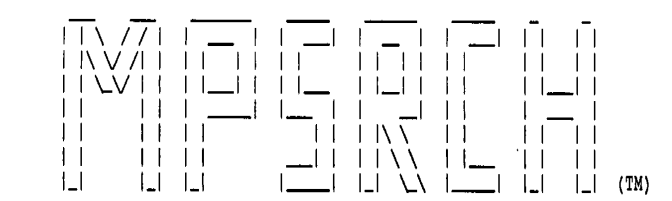
Query Match 24.6%; Score 239; DB 5; Length 683;
Best Local Similarity 29.5%; Pred. No. 7.66e-18;
Matches 36; Conservative 42; Mismatches 42; Indels 2; Gaps 2;

```
Db     81 KVSTPPHSLFQGFRNLDRLELDRCLIDTVPEGLFAGLGQLYSLIVKNAKITDFPREIFAH 140
          ::||  :: || :::|:|| |:|  ::  || || | ::|| | : :  |  :||  |
Qy     11 RISTIERGAFQDLKELERLRLNRNNLQLFPELLFLGTARLYRLDLSENQIQAIPRKAFRG 70

Db    141 VPNLMTLDLSGNRLR-IEPYSLRSLQNLIHLDVSDNDIGFLT-NTLISLTKLKVITMNNN 198
           ::: |:|  |::  ||  ::|:|::|  | :::|:|  |:  ::  : ||: : :::|
Qy     71 AVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRTFRLHSN 130

Db    199 KI 200
          ::
Qy    131 NL 132
```

Search completed: Fri May 28 09:32:09 1999
Job time : 52 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:34:00 1999; MasPar time 6.68 Seconds
331.067 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-13
Description: (1-104) from US09191647.pep
Perfect Score: 807
Sequence: 1 NNDDCVGHKCRHGAQCVDEV.....ITVNFVGKDSYVELASAKVR 104

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 27.273; Variance 127.182; scale 0.214

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	436	54.0	1534	30	W46966	Amino acid sequence o	1.73e-28
2	343	42.5	1480	5	R25079	Drosophila SLIT prote	2.51e-20
3	209	25.9	1193	19	W05835	Chick Serrate.	7.17e-09
4	207	25.7	612	28	W39256	Human partial mature	1.05e-08
5	207	25.7	737	28	W39257	Human membrane protei	1.05e-08
6	205	25.4	1872	36	W68510	Partial human Notch-3	1.54e-08
7	205	25.4	2321	36	W49698	Human Notch3 protein.	1.54e-08
8	203	25.2	1055	29	W44298	Human serrate 2 prote	2.26e-08
9	203	25.2	1212	29	W44299	Human serrate 2.	2.26e-08
10	203	25.2	1257	19	W05834	Human Serrate-2 (HJ2)	2.26e-08
11	200	24.8	685	37	W80813	Nucleotide sequence o	4.01e-08
12	190	23.5	1404	7	R38304	Sequence of a serrate	2.69e-07
13	188	23.3	722	21	W11720	M-Delta-1 polypeptide	3.93e-07
14	188	23.3	1036	25	W18351	Proliferation and dif	3.93e-07
15	188	23.3	1187	25	W18352	Proliferation and dif	3.93e-07
16	188	23.3	1218	29	W44301	Human serrate 1.	3.93e-07
17	188	23.3	1218	19	W05833	Human Serrate-1 (HJ1)	3.93e-07
18	188	23.3	1218	25	W18354	Proliferation and dif	3.93e-07
19	186	23.0	520	25	W18348	Proliferation and dif	5.73e-07
20	186	23.0	702	25	W18349	Proliferation and dif	5.73e-07
21	186	23.0	723	25	W18353	Proliferation and dif	5.73e-07
22	184	22.8	385	10	R56167	Neuroendocrine tumor	8.37e-07
23	183	22.7	1208	28	W40827	Human Jagged protein.	1.01e-06
24	178	22.1	727	21	W11719	C-Delta-1 polypeptide	2.59e-06
25	178	22.1	740	21	W00876	C-Delta-1 polypeptide	2.59e-06
26	172	21.3	660	21	W11725	H-Delta-1 polypeptide	7.98e-06
27	171	21.2	487	11	R60518	Cattle Factor-Xa.	9.62e-06
28	171	21.2	492	11	R60502	Serine protease for f	9.62e-06
29	171	21.2	833	6	R28960	Delta D11.	9.62e-06
30	164	20.3	383	10	R56166	Neuroendocrine tumor	3.55e-05
31	157	19.5	157	21	W11730	H-Delta-1 polypeptide	1.30e-04
32	155	19.2	467	29	W40283	Human Factor X protea	1.87e-04
33	152	18.8	415	7	R35761	Factor IX (IX).	3.26e-04
34	152	18.8	448	7	R37402	Factor X.	3.26e-04
35	152	18.8	448	35	W66092	Human factor X varian	3.26e-04
36	152	18.8	454	12	R67710	Human Factor-IX.	3.26e-04
37	152	18.8	461	3	P50302	Sequence of human fac	3.26e-04
38	152	18.8	461	29	W40284	Human Factor IX prote	3.26e-04
39	152	18.8	462	3	R10868	Recombinant human fac	3.26e-04
40	152	18.8	488	36	W76218	Human Factor X protei	3.26e-04
41	152	18.8	488	4	R22512	Mutated precursor of	3.26e-04
42	152	18.8	488	4	R22511	Human Factor Xai.	3.26e-04
43	152	18.8	488	36	W76216	Human Factor X protei	3.26e-04
44	152	18.8	488	36	W76219	Human Factor X protei	3.26e-04
45	152	18.8	488	36	W76217	Human Factor X protei	3.26e-04

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 54.0%; Score 436; DB 30; Length 1534;
Best Local Similarity 52.5%; Pred. No. 1.73e-28;
Matches 52; Conservative 24; Mismatches 22; Indels 1; Gaps 1;

Db 1083 nqddcrdhrcqngaqcmdevnsysclcaegysgqlceipphlpa-pkspcegtecqngan 1141
Qy 1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db 1142 cvdqgnrpvcqclpgfggpecekllsvnfvdrdtylqft 1180

```
Qy      61 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELA 99


RESULT    2
ID   R25079 standard; Protein; 1480 AA.
AC   R25079;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW   midline glial cells; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   peptide         1..36
FT                   /label= signal
FT   domain          73..294
FT                   /label= Flank_LRR_Flank_1
FT                   /note= "mediates adhesive events"
FT   domain          295..518
FT                   /label= Flank-LRR-Flank_2
FT                   /note= "mediates adhesive events"
FT   domain          519..714
FT                   /label= Flank_LRR_Flank_3
FT                   /note= "mediates adhesive events"
FT   domain          715..910
FT                   /label= Flank_LRR_Flank_4
FT                   /note= "mediates adhesive events"
FT   region          911..1150
FT                   /label= Tandem_EGF_like_repeats
FT                   /note= "involved in protein-protein interactions"
FT   region          1353..1393
FT                   /label= 7th_EGF_like_repeat
FT                   /note= "involved in receptor-ligand interactions"
FT   region          1394..1404
FT                   /label= alternative_splice_segment
FT                   /note= "developmentally regulated"
FT   region          1405..1480
FT                   /label= C-terminal_region
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   N-PSDB; Q25811.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 1; Page 84-89; 122pp; English.
CC   The SLIT protein is necessary for normal development of the midline
CC   of the CNS, partic. the midline glial cells, and for the
CC   concomitant formation of the commisural axon pathways. The process
CC   is dependent on the level of SLIT protein expression. It appears
CC   that SLIT protein is excreted by the midline glial cells where it
CC   is synthesised and is eventually associated with the surfaces of
CC   axons that traverse tham. The SLIT protein is tightly localised to
CC   the muscle attachment sites and to the sites of contact between
CC   adjacent pairs of cardioblasts as they coalesce to form the lumen
CC   of the larval heart. The SLIT protein defines a new set of
CC   molecules (TAGONS) which play a key role in axon outgrowth and
CC   pathfinding. SLIT can be used as a nerve regenerative in
CC   neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC   injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC   sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC   strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC   degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC   region-Flank" domains, the C-terminal region and part of the
CC   alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC   claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC   and at least 1 EGF-like repeat element from the SLIT protein.
CC   See also R29102.
SQ   Sequence   1480 AA;
```

```
  Query Match          42.5%;  Score 343;  DB 5;  Length 1480;
  Best Local Similarity  42.2%;  Pred. No. 2.51e-20;
  Matches   46;  Conservative   26;  Mismatches  31;  Indels   6;  Gaps   5;

Db   1064 niddcqnhmcqnggtcvdgindyqcrcpddytgkyceghnmismmypqtspcqnheckhg 1123
          | |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db   1124 v-cfqpnaqgsdylcrchpgytgkwceyltsisfvhnnsfveleplrtr 1171
           |:    | :   ||| ||::| :|| | :::|| ::|:||| : : |
Qy     59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    3
ID   W05835 standard; Protein; 1193 AA.
AC   W05835;
DT   28-JAN-1997  (first entry)
DE   Chick Serrate.
KW   C-Serrate; Notch; cell differentiation; cell fate; tissue repair;
KW   central nervous system; cancer; therapy; diagnosis.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          1..1041
FT                   /label= Extracellular_domain
FT   peptide         1..5
FT                   /label= Sig_peptide
FT                   /note= "lacks the N-terminal portion owing to
FT                   truncation of the encoding cDNA clone"
FT   domain          158..203
FT                   /label= DSL
FT                   /note= "region of homology with Drosophila Delta
FT                   and Serrate, predicted to mediate binding
FT                   with Notch"
FT   domain          208..837
FT                   /label= ELR
FT                   /note= "epidermal growth factor-like repeat domain"
FT   region          208..238
FT                   /label= ELR1
FT   region          239..274
FT                   /label= ELR2
FT   region          275..313
FT                   /label= ELR3
FT   region          314..351
FT                   /label= ELR4
FT   region          352..390
FT                   /label= ELR5
FT   region          391..427
FT                   /label= ELR6
FT   region          428..464
FT                   /label= ELR7
FT   region          465..502
FT                   /label= ELR8
FT   region          503..540
FT                   /label= ELR9
FT   region          541..606
FT                   /label= ELR10
FT   region          607..644
FT                   /label= ELR11
FT   region          655..682
FT                   /label= ELR12
FT   region          683..721
FT                   /label= ELR13
FT   region          722..759
FT                   /label= ELR14
FT   region          760..797
FT                   /label= ELR15
FT   region          798..837
FT                   /label= ELR16
FT   region          854..911
FT                   /label= Cysteine-rich_region
FT   domain          1042..1066
FT                   /label= Transmembrane_domain
```

```
FT   domain          1067..1193
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI   Lewis JH, Mann RS, Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40092.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Disclosure; Page 112-115; 161pp; English.
CC   Chicken Serrate (W05835), or C-Serrate, is a ligand for the zygotic
CC   neurogenic locus Notch and is believed to play a major role in
CC   determining cell fates in the central nervous system. Its amino
CC   acid sequence was deduced from a cDNA clone (T40092) obtd. from an
CC   optic explant cDNA library. C-Serrate is expressed in the central
CC   nervous system, cranial placodes, nephric mesoderm, vascular
CC   system, and limb bud mesenchyme.
SQ   Sequence  1193 AA;

  Query Match          25.9%;  Score 209;  DB 19;  Length 1193;
  Best Local Similarity  37.8%;  Pred. No. 7.17e-09;
  Matches   31;  Conservative   18;  Mismatches  26;  Indels   7;  Gaps   2;

Db    463 necasnpcmngghcqdeingfqclcpagfsgnlc-q------ldidycepnpcqngaqcf 515
          ::| :: | :|::| ||:||: |:|| |||| :| :     |:   |:   |||||||:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    516 nlamdyfcncpedyegkncshl 537
            :  :  | ||  : |  |  |
Qy     63 VVQQEPTCRCPPGFAGPRCEKL 84


RESULT   4
ID   W39256 standard; protein; 612 AA.
AC   W39256;
DT   19-MAY-1998  (first entry)
DE   Human partial mature membrane protein.
KW   Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW   brain; nervous tissue; cancer.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Protein        1..612
FT                  /note= "partial mature protein"
PN   J10036395-A.
PD   10-FEB-1998.
PF   24-JUL-1996; 194467.
PR   24-JUL-1996; JP-194467.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-174912/16.
PT   New human membrane protein - specifically expressed in brain and
PT   nervous tissue; used in diagnosis of diseases specific to these
PT   tissues and cancer
PS   Claim 1; Column 18-19; 26pp; Japanese.
CC   W39256 represents the partial mature amino acid sequence of a novel
CC   membrane protein which contains epidermal growth factor (EGF) motifs.
CC   The new membrane protein is expressed specifically in brain and nervous
CC   tissue. The protein and DNA can be used in the diagnosis of brain and
CC   nerve system specific diseases and cancer.
SQ   Sequence  612 AA;

  Query Match          25.7%;  Score 207;  DB 28;  Length 612;
  Best Local Similarity  42.0%;  Pred. No. 1.05e-08;
  Matches   34;  Conservative   13;  Mismatches  27;  Indels   7;  Gaps   3;

Db    370 cildpcrngatcisslsgftcqcpegyfgsacee---kv--d--pcasspcqnngtcyvd 422
          |:   ||:|| |:  ::|:|| ||:|: |  ||    :|  :  ||    ||| : | |
Qy      5 CVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIVV 64

Db    423 gvhftcncspgftgptcaqli 443
              || |:|||:|| |  ||
Qy     65 QQEPTCRCPPGFAGPRCEKLI 85


RESULT   5
ID   W39257 standard; protein; 737 AA.
AC   W39257;
DT   19-MAY-1998  (first entry)
DE   Human membrane protein.
KW   Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW   brain; nervous tissue; cancer; disease.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Peptide        1..26
FT                  /label= signal
FT   Protein        27..737
FT                  /label= membrane_protein
PN   J10036395-A.
PD   10-FEB-1998.
PF   24-JUL-1996; 194467.
PR   24-JUL-1996; JP-194467.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-174912/16.
DR   N-PSDB; V09641.
PT   New human membrane protein - specifically expressed in brain and
PT   nervous tissue; used in diagnosis of diseases specific to these
PT   tissues and cancer
PS   Claim 2; Pages 19-21; 26pp; Japanese.
CC   W39257 represents the amino acid sequence of a novel membrane protein
CC   which contains epidermal growth factor (EGF) motifs. The new membrane
CC   protein is expressed specifically in brain and nervous tissue. The
CC   protein and DNA can be used in the diagnosis of brain and nerve system
CC   specific diseases and cancer.
SQ   Sequence  737 AA;

  Query Match          25.7%;  Score 207;  DB 28;  Length 737;
  Best Local Similarity  42.0%;  Pred. No. 1.05e-08;
  Matches   34;  Conservative   13;  Mismatches  27;  Indels   7;  Gaps   3;

Db    396 cildpcrngatcisslsgftcqcpegyfgsacee---kv--d--pcasspcqnngtcyvd 448
          |:   ||:|| |:  ::|:|| ||:|: |  ||    :|  :  ||    ||| : | |
Qy      5 CVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIVV 64

Db    449 gvhftcncspgftgptcaqli 469
              || |:|||:|| |  ||
Qy     65 QQEPTCRCPPGFAGPRCEKLI 85


RESULT   6
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Misc_difference 328
FT                  /note= "encoded by NAN"
FT   Misc_difference 401
FT                  /note= "encoded by GNN"
FT   Misc_difference 403
FT                  /note= "encoded by GNC"
FT   Misc_difference 406
FT                  /note= "encoded by GNN"
FT   Misc_difference 409
FT                  /note= "encoded by NNT"
FT   Misc_difference 420
FT                  /note= "encoded by GNC"
FT   Misc_difference 706
```

```
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
SQ   Sequence   1872 AA;

  Query Match           25.4%;  Score 205;  DB 36;  Length 1872;
  Best Local Similarity  38.6%;  Pred. No. 1.54e-08;
  Matches    32;  Conservative   20;  Mismatches  23;  Indels   8;  Gaps   5;

Db    443 decastpcrngakcvdqpdgyecrcaegfegtlcdrn---vd-dcsp-dp--chhg-rcv 494
          |:| :  ||:|| |||: :|| | |::|| | :|::    |  : || |   |::| :|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    495 dgiasfscacapgytgtrcesqv 517
                :| |:||::| |||  :
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85


RESULT    7
ID   W49698 standard; Protein; 2321 AA.
AC   W49698;
DT   21-DEC-1998  (first entry)
DE   Human Notch3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy.
OS   Homo sapiens.
PN   FR2751986-A1.
PD   06-FEB-1998.
PF   16-APR-1997; 004680.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133138/13.
DR   N-PSDB; V57001.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1.1-1.8; 45pp; French.
CC   This sequence represents the human Notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of
CC   the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC   and leukoencephalopathy (CADASIL) type.  Blocking expression of a
CC   mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC   gene or protein can be used to treat CADASIL or related disorders.
SQ   Sequence   2321 AA;

  Query Match           25.4%;  Score 205;  DB 36;  Length 2321;
  Best Local Similarity  38.6%;  Pred. No. 1.54e-08;
  Matches    32;  Conservative   20;  Mismatches  23;  Indels   8;  Gaps   5;

Db    509 decastpcrngakcvdqpdgyecrcaegfegtlcdrn---vd-dcsp-dp--chhg-rcv 560
          |:| :  ||:|| |||: :|| | |::|| | :|::    |  : || |   |::| :|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    561 dgiasfscacapgytgtrcesqv 583
                :| |:||::| |||  :
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85


RESULT    8
ID   W44298 standard; Protein; 1055 AA.
AC   W44298;
DT   19-JUN-1998  (first entry)
DE   Human serrate 2 protein fragment.
KW   Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW   leukaemia; endothelial cell; tumour.
OS   Homo sapiens.
PN   WO9802458-A1.
PD   22-JAN-1998.
PF   11-JUL-1997; J02414.
PR   14-MAY-1997; JP-124063.
PR   16-JUL-1996; JP-186220.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A, Sakano S;
DR   WPI; 98-110528/10.
DR   N-PSDB; V15181.
PT   Human serrate-2 gene expression products  - used to regulate stem
PT   cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS   Claim 2; Page 57-62; 103pp; Japanese.
CC   The present sequence represents a human serrate 2 protein fragment. The
CC   present invention also describes a method for the preparation of the
CC   polypeptides, and antibodies binding to the polypeptide and its
CC   fragments. The polypeptide and its fragments expressed by the serrate-2
CC   gene can be used to inhibit stem (especially blood stem) cell
CC   differentiation and to inhibit endothelial cell growth. They may be
CC   incorporated in a cell culture media for culturing undifferentiated
CC   stem cells. They can also be used for treatment of neoplasms such as
CC   leukaemia. The antibodies can be used for the diagnosis of malignant
CC   tumours.
SQ   Sequence   1055 AA;

  Query Match           25.2%;  Score 203;  DB 29;  Length 1055;
  Best Local Similarity  42.7%;  Pred. No. 2.26e-08;
  Matches    35;  Conservative   15;  Mismatches  24;  Indels   8;  Gaps   2;

Db    435 nvndcrgq-cqhggtckdlvngyqcvcprgfggrhce-------lerdkcasspchsggl 486
          | :|| |: |:||: | | |||| |:||:||:|  ||       |:   |    |::|:
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    487 cedladgfhchcpqgfsgplce 508
          |  : :   |:|| ||:|| ||
Qy     61 CIVVQQEPTCRCPPGFAGPRCE 82


RESULT    9
ID   W44299 standard; Protein; 1212 AA.
AC   W44299;
DT   19-JUN-1998  (first entry)
DE   Human serrate 2.
KW   Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW   leukaemia; endothelial cell; tumour.
OS   Homo sapiens.
```

PN WO9802458-A1.
PD 22-JAN-1998.
PF 11-JUL-1997; J02414.
PR 14-MAY-1997; JP-124063.
PR 16-JUL-1996; JP-186220.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 98-110528/10.
DR N-PSDB; V15181.
PT Human serrate-2 gene expression products - used to regulate stem
PT cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS Claim 3; Page 62-68; 103pp; Japanese.
CC The present sequence represents human serrate 2. The present invention
CC also describes a method for the preparation of the polypeptides, and
CC antibodies binding to the polypeptide and its fragments. The polypeptide
CC and its fragments expressed by the serrate-2 gene can be used to inhibit
CC stem (especially blood stem) cell differentiation and to inhibit
CC endothelial cell growth. They may be incorporated in a cell culture
CC media for culturing undifferentiated stem cells. They can also be used
CC for treatment of neoplasms such as leukaemia. The antibodies can be used
CC for the diagnosis of malignant tumours.
SQ Sequence 1212 AA;

Query Match 25.2%; Score 203; DB 29; Length 1212;
Best Local Similarity 42.7%; Pred. No. 2.26e-08;
Matches 35; Conservative 15; Mismatches 24; Indels 8; Gaps 2;

Db 435 nvndcrgq-cqhggtckdlvngyqcvcprgfggrhce-------lerdkcasspchsggl 486
Qy 1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db 487 cedladgfhchcpqgfsgplce 508
Qy 61 CIVVQQEPTCRCPPGFAGPRCE 82

RESULT 10
ID W05834 standard; Protein; 1257 AA.
AC W05834;
DT 28-JAN-1997 (first entry)
DE Human Serrate-2 (HJ2).
KW Serrate-2; human jagged-2; HJ2; Notch; cell differentiation;
KW cell fate; central nervous system; cancer; tissue repair; therapy;
KW diagnosis; antibody.
OS Homo sapiens.

Key	Location/Qualifiers
FT domain	1..912
FT	/label= Extracellular_domain
FT	/note= "a deletion in the encoding cDNA clone results in loss of part of the Serrate-2 signal peptide and beginning of the DSL domain
FT domain	26..70
FT	/label= DSL
FT	/note= "region of homology with Drosophila Delta and Serrate, predicted to mediate binding with Notch"
FT domain	75..735
FT	/label= ELR
FT	/note= "epidermal growth factor-like repeat domain"
FT region	75..105
FT	/label= ELR1
FT region	106..140
FT	/label= ELR2
FT region	141..180
FT	/label= ELR3
FT region	181..218
FT	/label= ELR4
FT region	219..256
FT	/label= ELR5
FT region	257..294
FT	/label= ELR6
FT region	295..331
FT	/label= ELR7
FT region	332..369
FT	/label= ELR8
FT region	370..407
FT	/label= ELR9
FT region	408..435
FT	/label= Partial_ELR
FT region	436..469
FT	/label= Partial_ELR
FT region	470..507
FT	/label= ELR10
FT region	508..545
FT	/label= ELR11
FT region	546..584
FT	/label= ELR12
FT region	585..622
FT	/label= ELR13
FT region	623..660
FT	/label= ELR14
FT region	664..701
FT	/label= ELR15
FT region	702..718
FT	/label= Partial_ELR
FT region	719..735
FT	/label= Partial_ELR
FT domain	913..933
FT	/label= Transmembrane_domain
FT domain	934..1257
FT	/label= Intracellular_domain

PN WO9627610-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03172.
PR 07-MAR-1995; US-400159.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI Lewis JH, Mann RS, Myat AM;
DR WPI; 96-425379/42.
DR N-PSDB; W05834.
PT Vertebrate Serrate protein and related DNA - used to treat or
PT prevent malignancies characterised by increased Notch activity.
PS Claim 5; Page 104-107; 161pp; English.
CC Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC for the zygotic neurogenic locus Notch, and are believed to play a
CC major role in determining cell fates (differentiation) in the
CC central nervous system. Their amino acid sequences were deduced
CC from cDNA clones (see also T40090-91) isolated from human foetal
CC brain cDNA libraries. The proteins, antibodies raised to them,
CC and encoding nucleic acids can be used in the detection of
CC Serrate sequences and in the treatment of disorders of cell fate
CC or differentiation, partic. cancer, nervous system disorders
CC and in tissue repair or regeneration.
SQ Sequence 1257 AA;

Query Match 25.2%; Score 203; DB 19; Length 1257;
Best Local Similarity 42.7%; Pred. No. 2.26e-08;
Matches 35; Conservative 15; Mismatches 24; Indels 8; Gaps 2;

Db 291 nvndcrgq-cqhggtckdlvngyqcvcprgfggrhce-------lerdkcasspchsggl 342
Qy 1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db 343 cedladgfhchcpqgfsgplce 364
Qy 61 CIVVQQEPTCRCPPGFAGPRCE 82

RESULT 11
ID W80813 standard; Protein; 685 AA.
AC W80813;
DT 16-FEB-1999 (first entry)
DE Nucleotide sequence of the human Delta 3 protein.
KW Human; Delta 3 protein; agonist; tissue regeneration;

KW neurodegenerative disease; neurodifferentiative disorder;
KW neurodevelopmental disorder; peripheral neuropathy;
KW spinocerebella degeneration; antagonist; neoplastic disease;
KW hyperplastic disease; cancer; Waldenstroem's macroglobulemia;
KW fibroproliferative disorder; cerebravascular tissue; gene therapy;
KW antibody.
OS Homo sapiens.
PN WO9845434-A1.
PD 15-OCT-1998.
PF 06-APR-1998; U06775.
PR 11-JUN-1997; US-872855.
PR 04-APR-1997; US-832633.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Gearing DP, McCarthy SA;
DR WPI; 98-594482/50.
DR N-PSDB; V68523.
PT New isolated human Delta3 gene - used to develop products for
PT treating, e.g. nerve injury, neurodegenerative disorders, peripheral
PT neuropathies and spinocerebella degenerations
PS Claim 2; Fig 1; 160pp; English.
CC This is the amino acid sequence of the human Delta 3 protein
CC used in the method of the invention. The Delta3 gene is involved in
CC the growth and differentiation of cells. Delta3 agonists can be used
CC for promoting the tissue regeneration or repair needed to treat a
CC nerve injury, neurodegenerative disease, neurodifferentiative or
CC neurodevelopmental disorders including peripheral neuropathies and
CC spinocerebella degenerations. Delta3 antagonists can be used to treat
CC neoplastic or hyperplastic diseases, e.g. cancers, Waldenstroem's
CC macroglobulemia and fibroproliferative disorders, particularly of
CC cerebravascular tissue. The nucleic acids can also be used for gene
CC therapy. The products can also be used for antibody production,
CC detection, diagnosis and drug screening.
SQ Sequence 685 AA;

Query Match 24.8%; Score 200; DB 37; Length 685;
Best Local Similarity 41.5%; Pred. No. 4.01e-08;
Matches 34; Conservative 16; Mismatches 23; Indels 9; Gaps 4;

```
Db    327 ecdsnpcrnggsckdqedgyhclcppgyyglhcehst----l--s-cadspcfnggscre 379
          :| :: ||:|: | |: :|| |:|| |: || |||:     |  | | :  | ||: |
Qy      4 DCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIV 63

Db    380 rnqganyacecppnftgsncek 401
            | :  :| ||| |:|: |||
Qy     64 VQQEP--TCRCPPGFAGPRCEK 83
```

RESULT 12
ID R38304 standard; Protein; 1404 AA.
AC R38304;
DT 30-NOV-1993 (first entry)
DE Sequence of a serrate protein.
KW Serrate; toporythmic protein; family.
OS Drosophila melanogaster.
PN WO9312141-A.
PD 24-JUN-1993.
PF 11-DEC-1991; U09240.
PR 11-DEC-1991; WO-U09240.
PA (UYYA) UNIV YALE.
PI Artavanis-tsakonas S, Fleming RJ;
DR WPI; 93-214095/26.
DR N-PSDB; Q43910.
PT Purified serrate protein, nucleic acid and antibodies - used in
PT the study and manipulation of differentiation and other
PT physiological processes
PS Claim 4; Pages 74-80; 119pp; English.
CC Two Drosophila ganomic phage libraries were screened and recombinant
CC clones were isolated. The cDNAs in lambda gt10 were isolated from an
CC early pupal library. The C1 cDNA was isolated from an early pupal
CC library. Subsequently the C3 cDNA was isolated using the 5' 700 bp
CC terminal fragment of the C1 cDNA as probe. The complete 5561bp
CC sequence of DNA of the Drosophila Serrate protein was derived from
CC C1 and C3 cDNAs (Q43910). The deduced protein product appears to be
CC a transmembrane protein. AAs 51-80 represent the likely signal
CC peptide; aas 542-564 represent potential membrane associated region;
CC aas 1221-1245 represent the putative transmembrane domain.
SQ Sequence 1404 AA;

Query Match 23.5%; Score 190; DB 7; Length 1404;
Best Local Similarity 41.3%; Pred. No. 2.69e-07;
Matches 33; Conservative 15; Mismatches 24; Indels 8; Gaps 6;

```
Db    609 ddcvgq-crngatcidlvndyrcacasgftgrdce--td-i--d-e-catspcrnggecv 660
          |||||: ||:|| |:| || | | |: ||:|  ||     :  :   |    |:||::|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    661 dmvgkfncicplgysgslce 680
            :    | || |::|: ||
Qy     63 VVQQEPTCRCPPGFAGPRCE 82
```

RESULT 13
ID W11720 standard; Protein; 722 AA.
AC W11720;
DT 28-APR-1997 (first entry)
DE M-Delta-1 polypeptide.
KW M-Delta-1; cell proliferation; nervous system disorder;
KW tissue regeneration; Notch; cervix cancer; breast cancer;
KW lung cancer; colon cancer; melanoma; seminoma;
KW neurogenesis; therapy.
OS Mus sp.
PN WO9701571-A1.
PD 16-JAN-1997.
PF 28-JUN-1996; U11178.
PR 28-JUN-1995; US-000589.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique D, Ish-Horowicz D;
PI Lewis J;
DR WPI; 97-100159/09.
DR N-PSDB; T58899.
PT New vertebrate Delta protein, DNA and antibodies - for treating and
PT preventing cancer, nervous system disorders and for tissue
PT regeneration
PS Claim 4; Fig 8; 135pp; English.
CC M-delta-1 polypeptide (W11720) is the mouse homologue of Drosophila
CC Delta, a protein that binds to Notch protein. It is expressed
CC primarily in presomitic mesoderm, the central and peripheral
CC nervous systems, and kidney. Chick (W11719) and human (W11721-
CC 38) Delta-1 polypeptides have also been identified. Delta-1
CC proteins can be used to treat or prevent disorders characterised by
CC increased Notch activity, such as cervical, breast, lung or colon
CC cancer, melanoma or seminoma, as well as nervous system disorders,
CC and to promote tissue regeneration and repair.
SQ Sequence 722 AA;

Query Match 23.3%; Score 188; DB 21; Length 722;
Best Local Similarity 38.8%; Pred. No. 3.93e-07;
Matches 33; Conservative 17; Mismatches 28; Indels 7; Gaps 3;

```
Db    442 nvddcasspcanggtcrdsvndfsctcppgytgknc--sap-v----srcehapchngat 494
          | ||| :  | :|: | | || ::| || |::|  |  ::| |    | |::  |:|||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    495 chqrgqrymcecaqgyggpncqfll 519
            |   |  | |: |::|| |: |:
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLI 85
```

RESULT 14
ID W18351 standard; protein; 1036 AA.
AC W18351;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;
KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;

KW immunosuppression.
OS Homo sapiens.
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 5; Page 66-71; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1036 AA;

Query Match 23.3%; Score 188; DB 25; Length 1036;
Best Local Similarity 36.6%; Pred. No. 3.93e-07;
Matches 30; Conservative 17; Mismatches 28; Indels 7; Gaps 2;

Db 458 decasnpclngghcqneinrfqclcptgfsgnlc-q------ldidycepnpcqngaqcy 510
Qy 3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db 511 nrasdyfckcpedyegkncshl 532
Qy 63 VVQQEPTCRCPPGFAGPRCEKL 84

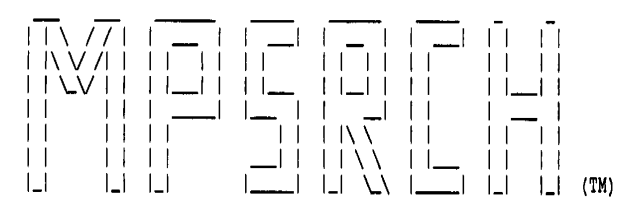
RESULT 15
ID W18352 standard; protein; 1187 AA.
AC W18352;
DT 11-FEB-1998 (first entry)
DE Proliferation and differentiation suppression polypeptide.
KW Proliferation; differentiation; suppression; human; delta-1;
KW serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW immunosuppression.
OS Homo sapiens.
PN WO9719172-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; J03356.
PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 6; Page 71-76; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
SQ Sequence 1187 AA;

Query Match 23.3%; Score 188; DB 25; Length 1187;
Best Local Similarity 36.6%; Pred. No. 3.93e-07;
Matches 30; Conservative 17; Mismatches 28; Indels 7; Gaps 2;

Db 458 decasnpclngghcqneinrfqclcptgfsgnlc-q------ldidycepnpcqngaqcy 510
Qy 3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db 511 nrasdyfckcpedyegkncshl 532
Qy 63 VVQQEPTCRCPPGFAGPRCEKL 84

Search completed: Fri May 28 09:34:58 1999
Job time : 58 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

ssrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:35:16 1999; MasPar time 7.12 Seconds
585.478 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-13
Description: (1-104) from US09191647.pep
Perfect Score: 807
Sequence: 1 NNDDCVGHKCRHGAQCVDEV.....ITVNFVGKDSYVELASAKVR 104

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 36.296; Variance 59.185; scale 0.613

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	343	42.5	530	2	A31640	epidermal growth fact	1.29e-57
2	343	42.5	1469	2	B36665	slit protein 2 precur	1.29e-57
3	343	42.5	1480	2	A36665	slit protein 1 precur	1.29e-57
4	215	26.6	2555	2	A40043	notch protein homolog	8.30e-28
5	214	26.5	1064	2	A40136	fibropellin Ia - sea	1.39e-27
6	209	25.9	2703	2	A24420	notch protein - fruit	1.81e-26
7	205	25.4	293	2	B26637	neurogenic repetitive	1.40e-25
8	205	25.4	2139	2	A35672	crumbs protein - frui	1.40e-25
9	205	25.4	2321	2	S78549	notch3 protein - huma	1.40e-25
10	205	25.4	2524	2	A35844	Xotch protein - Afric	1.40e-25
11	202	25.0	2318	2	S45306	notch 3 protein - mou	6.47e-25
12	198	24.5	861	2	A48825	Notch homolog Motch p	4.94e-24
13	198	24.5	2531	2	A46019	gene Notch-1 protein	4.94e-24
14	198	24.5	2531	2	S18188	notch protein homolog	4.94e-24
15	197	24.4	387	2	B49175	Motch A protein - mou	8.20e-24
16	196	24.3	1203	2	A49175	Motch B protein - mou	1.36e-23
17	195	24.2	1429	2	S06434	homeotic protein lin-	2.25e-23
18	195	24.2	2471	2	A49128	cell-fate determining	2.25e-23
19	193	23.9	570	2	A48836	fibropellin C precurs	6.19e-23
20	192	23.8	2437	2	S42612	transmembrane protein	1.02e-22
21	190	23.5	1404	2	A36666	serrate protein precu	2.80e-22
22	190	23.5	1408	2	S16148	gene serrate protein	2.80e-22
23	188	23.3	722	2	I48324	DELTA-like 1 - mouse	7.64e-22
24	188	23.3	1220	2	A56136	jagged protein precur	7.64e-22
25	184	22.8	385	2	A54785	preadipocyte factor 1	5.65e-21
26	184	22.8	385	2	S53718	homeotic protein dlk	5.65e-21
27	178	22.1	728	2	I50719	C-Delta-1 - chicken	1.12e-19
28	175	21.7	4391	2	A38096	perlecan precursor -	4.91e-19
29	171	21.2	492	1	EXBO	coagulation factor Xa	3.51e-18
30	171	21.2	832	2	A31246	neurogenic protein De	3.51e-18
31	171	21.2	833	2	S19087	gene Delta protein pr	3.51e-18
32	171	21.2	880	2	S00670	gene Delta protein pr	3.51e-18
33	171	21.2	4543	1	A53102	alpha-2-macroglobulin	3.51e-18
34	165	20.4	475	1	EXCH	coagulation factor Xa	6.57e-17
35	164	20.3	200	2	A26637	neurogenic repetitive	1.07e-16
36	164	20.3	259	2	S48713	fetal antigen 1 - hum	1.07e-16
37	164	20.3	260	2	A44549	fetal antigen 1 homeo	1.07e-16
38	164	20.3	383	2	B45484	delta-like dlk homeot	1.07e-16
39	164	20.3	383	2	S53716	homeotic protein dlk	1.07e-16
40	163	20.2	3051	2	S42373	hypothetical protein	1.73e-16
41	162	20.1	1295	2	A32901	glp1 protein precurso	2.81e-16
42	157	19.5	482	1	EXRT	coagulation factor Xa	3.13e-15
43	154	19.1	443	2	I46932	coagulation factor VI	1.31e-14
44	152	18.8	427	2	S74211	PAS-6/7 protein precu	3.40e-14
45	152	18.8	488	1	EXHU	coagulation factor Xa	3.40e-14

ALIGNMENTS

RESULT 1
ENTRY A31640 #type fragment
TITLE epidermal growth factor-like protein slit - fruit fly (Drosophila melanogaster) (fragment)
ORGANISM #formal_name Drosophila melanogaster
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 14-Aug-1998
ACCESSIONS A31640
REFERENCE A31640
#authors Rothberg, J.M.; Hartley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
#journal Cell (1988) 55:1047-1059
#title slit: An EGF-homologous locus of D. melanogaster involved in the development of the embryonic central nervous system.
#cross-references MUID:89077533
#accession A31640
##molecule_type DNA
##residues 1-530 ##label ROT
##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
#gene FlyBase:sli
##cross-references FlyBase:FBgn0003425
#introns 470/3
CLASSIFICATION #superfamily EGF homology
KEYWORDS growth factor
FEATURE
148-181 #domain EGF homology #label EGF
SUMMARY #length 530 #checksum 6330

Query Match 42.5%; Score 343; DB 2; Length 530;
Best Local Similarity 42.2%; Pred. No. 1.29e-57;
Matches 46; Conservative 26; Mismatches 31; Indels 6; Gaps 5;

```
Db   184 NIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHG 243
             |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy     1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db   244 V-CFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTR 291
           |:   | :  ||| ||::| :|| | :::|| ::|:||| : : |
Qy    59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104
```

RESULT 2
ENTRY B36665 #type complete
TITLE slit protein 2 precursor - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster

```
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  16-Dec-1998
ACCESSIONS      B36665
REFERENCE       A36665
   #authors     Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
   #journal     Genes Dev. (1990) 4:2169-2187
   #title       slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
   #cross-references MUID:91099665
   #accession   B36665
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-1469 ##label ROT
      ##cross-references GB:X53959
GENETICS
   #gene        FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
FEATURE
   66-91            #domain proteoglycan amino-terminal homology #label
                      PAH1\
   101-124          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   125-148          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   149-172          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   173-196          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   197-220          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   228-272          #domain proteoglycan carboxyl-terminal homology #label
                      PCS1\
   288-313          #domain proteoglycan amino-terminal homology #label
                      PAH2\
   323-346          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   347-370          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   371-394          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   395-418          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   419-442          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   450-494          #domain proteoglycan carboxyl-terminal homology #label
                      PCS2\
   512-537          #domain proteoglycan amino-terminal homology #label
                      PAH3\
   547-571          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   572-595          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12\
   596-619          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR13\
   620-643          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR14\
   651-695          #domain proteoglycan carboxyl-terminal homology #label
                      PCS3\
   708-733          #domain proteoglycan amino-terminal homology #label
                      PAH4\
   743-766          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR15\
   767-790          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR16\
   846-890          #domain proteoglycan carboxyl-terminal homology #label
                      PCS4\
   1028-1061        #domain EGF homology #label EGF
SUMMARY         #length 1469  #molecular-weight 164695  #checksum 8361
```

```
  Query Match          42.5%;  Score 343;  DB 2;  Length 1469;
  Best Local Similarity  42.2%;  Pred. No. 1.29e-57;
  Matches    46;  Conservative   26;  Mismatches  31;  Indels   6;  Gaps   5;

Db   1064 NIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHG 1123
          | |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db   1124 V-CFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTR 1171
           |:    | :  ||| ||::| :|| | :::|| ::|:||| : : |
Qy     59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    3
ENTRY           A36665     #type complete
TITLE           slit protein 1 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  24-Sep-1998
ACCESSIONS      A36665; S13523
REFERENCE       A36665
   #authors     Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
   #journal     Genes Dev. (1990) 4:2169-2187
   #title       slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
   #cross-references MUID:91099665
   #accession   A36665
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-1480 ##label ROT
      ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
   #gene        FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
KEYWORDS        alternative splicing
FEATURE
   66-91            #domain proteoglycan amino-terminal homology #label
                      PAH1\
   101-124          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
   125-148          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
   149-172          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
   173-196          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
   197-220          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
   228-272          #domain proteoglycan carboxyl-terminal homology #label
                      PCS1\
   288-313          #domain proteoglycan amino-terminal homology #label
                      PAH2\
   323-346          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
   347-370          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
   371-394          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
   395-418          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
   419-442          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR10\
   450-494          #domain proteoglycan carboxyl-terminal homology #label
                      PCS2\
   512-537          #domain proteoglycan amino-terminal homology #label
                      PAH3\
```

```
   547-571          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR11\
   572-595          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR12\
   596-619          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR13\
   620-643          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR14\
   651-695          #domain proteoglycan carboxyl-terminal homology #label
                      PCS3\
   708-733          #domain proteoglycan amino-terminal homology #label
                      PAH4\
   743-766          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR15\
   767-790          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR16\
   791-814          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR17\
   815-838          #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LR18\
   846-890          #domain proteoglycan carboxyl-terminal homology #label
                      PCS4\
   1028-1061        #domain EGF homology #label EGF
SUMMARY          #length 1480  #molecular-weight 165751  #checksum 900

  Query Match           42.5%;  Score 343;  DB 2;  Length 1480;
  Best Local Similarity  42.2%;  Pred. No. 1.29e-57;
  Matches   46;  Conservative   26;  Mismatches  31;  Indels   6;  Gaps   5;

Db   1064 NIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHG 1123
          | |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db   1124 V-CFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTR 1171
           |:    | :   ||| ||::| :|| | :::|| ::|:||| : : |
Qy     59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    4
ENTRY            A40043     #type complete
TITLE            notch protein homolog TAN-1 precursor - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
                   14-Aug-1998
ACCESSIONS       A40043
REFERENCE        A40043
   #authors      Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds,
                   T.C.; Smith, S.D.; Sklar, J.
   #journal      Cell (1991) 66:649-661
   #title        TAN-1, the human homolog of the Drosophila Notch gene, is
                   broken by chromosomal translocations in T lymphoblastic
                   neoplasms.
   #cross-references MUID:91347367
   #accession    A40043
      ##status        preliminary; nucleic acid sequence not shown; not
                       compared with conceptual translation
      ##molecule_type mRNA
      ##residues      1-2555 ##label ELL
      ##cross-references GB:M73980
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   1149-1180        #domain EGF homology #label EGF\
   1927-1959        #domain ankyrin repeat homology #label AN1\
   1960-1992        #domain ankyrin repeat homology #label AN2\
   1994-2026        #domain ankyrin repeat homology #label AN3\
   2027-2059        #domain ankyrin repeat homology #label AN4\
   2060-2092        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2555  #molecular-weight 272337  #checksum 463

  Query Match           26.6%;  Score 215;  DB 2;  Length 2555;
  Best Local Similarity  41.0%;  Pred. No. 8.30e-28;
  Matches   34;  Conservative   18;  Mismatches  24;  Indels   7;  Gaps   4;

Db    947 NECASDPCRNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCT-ESS-C--F---NGGTCV 999
          ::| :  ||:|| | | |::||| || ||||: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINSFTCLCPPGFTGSYCQHVV 1022
               || |||||:|: |: ::
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85


RESULT    5
ENTRY            A40136     #type complete
TITLE            fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
ALTERNATE_NAMES  epidermal growth factor homolog precursor
CONTAINS         alternatively spliced fibropellin Ib (EGFI)
ORGANISM         #formal_name Strongylocentrotus purpuratus #common_name
                   purple urchin
DATE             13-May-1992 #sequence_revision 17-Sep-1997 #text_change
                   07-Aug-1998
ACCESSIONS       A40136; B40136; C40136; A29316; A43131
REFERENCE        A40136
   #authors      Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff,
                   R.A.
   #journal      J. Mol. Evol. (1989) 29:314-327
   #title        Structural analysis of the uEGF gene in the sea urchin
                   Strongylocentrotus purpuratus reveals more similarity to
                   vertebrate than to invertebrate genes with EGF-like
                   repeats.
   #cross-references MUID:90112459
   #accession    A40136
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-114 ##label DEL
      ##cross-references GB:X17530; NID:g10225; PID:g667061
   #accession    B40136
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues      181-251,329-370,'R',372-408,'RA',411-441 ##label DE2
   #accession    C40136
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues      'K',747-821,898-978 ##label DE3
REFERENCE        A29316
   #authors      Hursh, D.A.; Andrews, M.E.; Raff, R.A.
   #journal      Science (1987) 237:1487-1490
   #title        A sea urchin gene encodes a polypeptide homologous to
                   epidermal growth factor.
   #cross-references MUID:87319677
   #accession    A29316
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      'S',280-481,786-1064 ##label HUR
      ##cross-references GB:M17421; NID:g161474; PID:g552260
REFERENCE        A43131
   #authors      Hunt, L.T.; Barker, W.C.
   #journal      FASEB J. (1989) 3:1760-1764
   #title        Avidin-like domain in an epidermal growth factor homolog from
                   a sea urchin.
   #cross-references MUID:89196806
   #contents     annotation
COMMENT      EGF homology repeats 10-17 are spliced out in the short form
               (fibropellin Ib).
CLASSIFICATION   #superfamily C1r/C1s repeat homology; EGF homology
FEATURE
   1-19             #domain signal sequence #status predicted #label SIG\
   20-1064          #product fibropellin I #status predicted #label FIB\
   23-54            #domain EGF homology #label EG01\
   57-175           #domain C1r/C1s repeat homology #label CSR\
   180-211          #domain EGF homology #label EG02\
   218-249          #domain EGF homology #label EG03\
   256-287          #domain EGF homology #label EG04\
   294-325          #domain EGF homology #label EG05\
   332-363          #domain EGF homology #label EG06\
```

```
   370-401          #domain EGF homology #label EG07\
   408-439          #domain EGF homology #label EG08\
   446-477          #domain EGF homology #label EG09\
   484-515          #domain EGF homology #label EG10\
   522-553          #domain EGF homology #label EG11\
   560-591          #domain EGF homology #label EG12\
   598-629          #domain EGF homology #label EG13\
   636-667          #domain EGF homology #label EG14\
   674-705          #domain EGF homology #label EG15\
   712-743          #domain EGF homology #label EG16\
   750-781          #domain EGF homology #label EG17\
   788-819          #domain EGF homology #label EG18\
   826-857          #domain EGF homology #label EG19\
   864-895          #domain EGF homology #label EG20\
   902-933          #domain EGF homology #label EG21\
   936-1064         #region avidin-like\
   23-34,28-43,45-54,
   62-88,180-191,
   185-200,202-211,
   218-229,223-238,
   240-249,256-267,
   261-276,278-287,
   294-305,299-314,
   316-325,332-343,
   337-352,354-363,
   370-381,375-390,
   392-401,408-419,
   413-428,430-439,
   446-457,451-466,
   468-477,484-495   #disulfide_bonds #status predicted\
   489-504,506-515,
   522-533,527-542,
   544-553,560-571,
   565-580,582-591,
   598-609,603-618,
   620-629,636-647,
   641-656,658-667,
   674-685,679-694,
   696-705,712-723,
   717-732,734-743,
   750-761,755-770,
   772-781,788-799,
   793-808,810-819,
   826-837,831-846,
   848-857,864-875,
   869-884,886-895,
   902-913,907-922,
   924-933          #disulfide_bonds #status predicted
SUMMARY         #length 1064  #molecular-weight 112072  #checksum 303

 Query Match          26.5%;  Score 214;  DB 2;  Length 1064;
 Best Local Similarity  42.7%;  Pred. No. 1.39e-27;
 Matches   35;  Conservative   13;  Mismatches  27;  Indels   7;  Gaps   2;

Db    822 NIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCE-----ISLDA--CRSMPCQNGAT 874
          | |:| :  | :|: ||| |||: | ||  :|| :||     : |:: |    |||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    875 CVNVGADYVCECVPGYAGQNCE 896
          |: |  :  | | ||:||  ||
Qy     61 CIVVQQEPTCRCPPGFAGPRCE 82


RESULT    6
ENTRY           A24420     #type complete
TITLE           notch protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  neurogenic repetitive locus protein
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
                  07-Aug-1998
ACCESSIONS      A24420; A24768; S09358; A05267
REFERENCE       A24420
  #authors      Kidd, S.; Kelley, M.R.; Young, M.W.
  #journal      Mol. Cell. Biol. (1986) 6:3094-3108
  #cross-references MUID:87064624
  #accession    A24420
    ##molecule_type DNA
    ##residues     1-2703 ##label KID
    ##cross-references GB:K03508; NID:g157991; PID:g157993
REFERENCE       A24768
  #authors      Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
  #journal      Cell (1985) 43:567-581
  #cross-references MUID:86079539
  #accession    A24768
    ##molecule_type mRNA
    ##residues     1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',
                   268-872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',
                   2258-2264,'V',2266-2406,'R',2408-2444,'L',2446-2703
                   ##label WHA1
    ##note        the authors translated the codon ATC for residue 49 as
                   Thr, ATT for residue 2044 as Arg, GTA for residue 2265
                   as Ala, CGC for residue 2407 as His, and CTT for
                   residue 2445 as Arg
REFERENCE       S09358
  #authors      Tautz, D.
  #journal      Nucleic Acids Res. (1989) 17:6463-6471
  #title        Hypervariability of simple sequences as a general source for
                 polymorphic DNA markers.
  #cross-references MUID:89385974
  #accession    S09358
    ##molecule_type DNA
    ##residues     2505-2551,'QQQQ',2552-2576,'E',2578-2604 ##label TAU
REFERENCE       A05267
  #authors      Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.;
                 Artavanis-Tsakonas, S.
  #journal      Cell (1985) 40:55-62
  #title        opa: a novel family of transcribed repeats shared by the
                 Notch locus and other developmentally regulated loci in D.
                 melanogaster.
  #cross-references MUID:85099329
  #accession    A05267
    ##molecule_type DNA
    ##residues     2504-2576,'E',2578-2611 ##label WHA2
GENETICS
  #gene         notch; opa
    ##cross-references FlyBase:FBgn0004647
  #map_position 8.96-9.36
  #introns      53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
CLASSIFICATION  #superfamily notch protein; ankyrin repeat homology; EGF
                  homology
KEYWORDS        differentiation; tandem repeat; transmembrane protein
FEATURE
   27-43            #domain transmembrane #status predicted #label TMM1\
   568-599          #domain EGF homology #label EGF\
   1746-1762        #domain transmembrane #status predicted #label TMM2\
   1950-1982        #domain ankyrin repeat homology #label AN1\
   1983-2015        #domain ankyrin repeat homology #label AN2\
   1988-2004        #domain transmembrane #status predicted #label TMM3\
   2017-2049        #domain ankyrin repeat homology #label AN3\
   2050-2082        #domain ankyrin repeat homology #label AN4\
   2083-2115        #domain ankyrin repeat homology #label AN5\
   2538-2568        #region glutamine-rich\
   2538-2568        #domain neurogenic repetitive element #status predicted
                      #label OPA
SUMMARY         #length 2703  #molecular-weight 288876  #checksum 6404

  Query Match          25.9%;  Score 209;  DB 2;  Length 2703;
  Best Local Similarity  40.7%;  Pred. No. 1.81e-26;
  Matches   33;  Conservative   14;  Mismatches  30;  Indels   4;  Gaps   2;

Db    293 NYDDCLGHLCQNGGTCIDGISDYTCRCPPNFTGRFCQDD---VD-ECAQRDHPVCQNGAT 348
          | |||:|| |::|: |:| :: ||| ||  |:| ||:     | : :  |:  |||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    349 CTNTHGSYSCICVNGWAGLDC 369
          |  :  :| |  | ||  |
```

```
Qy     61 CIVVQQEPTCRCPPGFAGPRC 81


RESULT    7
ENTRY           B26637     #type fragment
TITLE           neurogenic repetitive locus 95F protein - fruit fly
                  (Drosophila melanogaster) (fragment)
ORGANISM        #formal_name Drosophila melanogaster
DATE            16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
                  14-Aug-1998
ACCESSIONS      B26637
REFERENCE       A91081
   #authors     Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigel,
                  D.; Vaessin, H.; Campos-Ortega, J.A.
   #journal     EMBO J. (1987) 6:761-766
   #title       EGF homologous sequences encoded in the genome of Drosophila
                  melanogaster, and their relation to neurogenic genes.
   #cross-references MUID:87218537
   #accession   B26637
      ##molecule_type mRNA
      ##residues      1-293 ##label KNU
      ##cross-references GB:X05144; NID:g7519; PID:g929536
GENETICS
   #gene        FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
   216-252           #domain EGF homology #label EGF
SUMMARY         #length 293  #checksum 3413

  Query Match           25.4%;  Score 205;  DB 2;  Length 293;
  Best Local Similarity  34.9%;  Pred. No. 1.40e-25;
  Matches    29;  Conservative   14;  Mismatches  37;  Indels   3;  Gaps   3;

Db    173 NIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTC-ENEPCRNGSTCQNGF 231
          | |:|  : |::|: | | : :| | ||: : |  |:    |   :  || :   ||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPC-DQYECQNGA 59

Db    232 N-ASTGNNFTCTCVPGFEGPLCD 253
              : || | ||| || |:
Qy     60 QCIVVQQEPTCRCPPGFAGPRCE 82


RESULT    8
ENTRY           A35672     #type complete
TITLE           crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A35672
REFERENCE       A35672
   #authors     Tepass, U.; Theres, C.; Knust, E.
   #journal     Cell (1990) 61:787-799
   #title       crumbs encodes an EGF-like protein expressed on apical
                  membranes of Drosophila epithelial cells and required for
                  organization of epithelia.
   #cross-references MUID:90263104
   #accession   A35672
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-2139 ##label TEP
      ##cross-references GB:M33753
      ##note         the authors translated the codon GGC for residue 1928 as
                      Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene        FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
   691-722           #domain EGF homology #label EGF
SUMMARY         #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           25.4%;  Score 205;  DB 2;  Length 2139;
  Best Local Similarity  34.9%;  Pred. No. 1.40e-25;
  Matches    29;  Conservative   14;  Mismatches  37;  Indels   3;  Gaps   3;

Db   1835 NIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTC-ENEPCRNGSTCQNGF 1893
          | |:|  : |::|: | | : :| | ||: : |  |:    |   :  || :   ||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPC-DQYECQNGA 59

Db   1894 N-ASTGNNFTCTCVPGFEGPLCD 1915
              : || | ||| || |:
Qy     60 QCIVVQQEPTCRCPPGFAGPRCE 82


RESULT    9
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
   #authors     Joutel, A.; Tournier-Lasserve, E.
   #submission  submitted to the EMBL Data Library, April 1997
   #accession   S78549
      ##molecule_type mRNA
      ##residues      1-2321 ##label JOU1
      ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE       S71825
   #authors     Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
   #journal     Nature (1996) 383:707-710
   #title       Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
   #cross-references MUID:97032728
   #accession   S71825
      ##status       nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues      67-113;138-194;268-333,'G',335-346;536-613;716-765;
                       1240-1279;1815-1888 ##label JOU2
      ##cross-references EMBL:U97669
GENETICS
   #gene        notch3
   #map_position 19p13.1
FUNCTION
   #description  may be involved in pathogenesis of CADASIL, causing a type of
                   stroke and dementia
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        tandem repeat; transmembrane protein
FEATURE
   318-349           #domain EGF homology #label EGF\
   1838-1870         #domain ankyrin repeat homology #label AN1\
   1871-1903         #domain ankyrin repeat homology #label AN2\
   1905-1937         #domain ankyrin repeat homology #label AN3\
   1938-1970         #domain ankyrin repeat homology #label AN4\
   1971-2003         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2321  #molecular-weight 243657  #checksum 3337

  Query Match           25.4%;  Score 205;  DB 2;  Length 2321;
  Best Local Similarity  38.6%;  Pred. No. 1.40e-25;
  Matches    32;  Conservative   20;  Mismatches  23;  Indels   8;  Gaps   5;

Db    509 DECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCDRN---VD-DCSP-DP--CHHG-RCV 560
          |:| :  ||:|| |||: :|| | |::|| | :|::    |  : || |   |::| :|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    561 DGIASFSCACAPGYTGTRCESQV 583
              :| |:||::| |||  :
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85
```

```
RESULT   10
ENTRY            A35844     #type complete
TITLE            Xotch protein - African clawed frog
ORGANISM         #formal_name Xenopus laevis #common_name African clawed frog
DATE             12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change
                   14-Aug-1998
ACCESSIONS       A35844
REFERENCE        A35844
   #authors      Coffman, C.; Harris, W.; Kintner, C.
   #journal      Science (1990) 249:1438-1441
   #title        Xotch, the Xenopus homolog of Drosophila notch.
   #cross-references MUID:90385285
   #accession    A35844
      ##status       preliminary; nucleic acid sequence not shown; not
                      compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-2524 ##label COF
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
   WORDS         transmembrane protein
   ATURE
   222-254           #domain EGF homology #label EGF\
   1924-1956         #domain ankyrin repeat homology #label AN1\
   1957-1989         #domain ankyrin repeat homology #label AN2\
   1991-2023         #domain ankyrin repeat homology #label AN3\
   2024-2056         #domain ankyrin repeat homology #label AN4\
   2057-2089         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2524  #molecular-weight 274931  #checksum 9441

  Query Match           25.4%;  Score 205;  DB 2;  Length 2524;
  Best Local Similarity  42.5%;  Pred. No. 1.40e-25;
  Matches   34;  Conservative   16;  Mismatches  23;  Indels   7;  Gaps   4;

Db    947 NECASNPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCT-ESS-C--F---NGGTCI 999
          ::| :: |::|| | | ||:||| |  ||||: ||   |    ::| |  :   ||: ||
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINTFTCQCPPGFTGSYCQ 1019
                ||:|||||:|: |:
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT   11
ENTRY            S45306     #type complete
TITLE            notch 3 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   10-Jul-1998
   CESSIONS      S45306
REFERENCE        S45306
   #authors      Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal      Mech. Dev. (1994) 46:123-136
   #title        The novel Notch homologue mouse Notch 3 lacks specific
                   epidermal growth factor-repeats and is expressed in
                   proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession    S45306
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1839-1871         #domain ankyrin repeat homology #label AN1\
   1872-1904         #domain ankyrin repeat homology #label AN2\
   1906-1938         #domain ankyrin repeat homology #label AN3\
   1939-1971         #domain ankyrin repeat homology #label AN4\
   1972-2004         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           25.0%;  Score 202;  DB 2;  Length 2318;
  Best Local Similarity  39.8%;  Pred. No. 6.47e-25;
  Matches   33;  Conservative   19;  Mismatches  23;  Indels   8;  Gaps   5;

Db    510 DECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCERN---VD-DCSP-DP--CHHG-RCV 561
          |:| :  ||:|| |||: :|| | |::|| | :||:    |  : || |   |::| :|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    562 DGIASFSCACAPGYTGIRCESQV 584
                :| |:||::| |||  :
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85


RESULT   12
ENTRY            A48825     #type fragment
TITLE            Notch homolog Motch protein - mouse (fragment)
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
                   14-Aug-1998
ACCESSIONS       A48825
REFERENCE        A48825
   #authors      Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.;
                   Rossant, J.
   #journal      Dev. Biol. (1992) 154:377-387
   #title        Expression analysis of a Notch homologue in the mouse embryo.
   #cross-references MUID:93050801
   #accession    A48825
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-861 ##label REA
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:119144)
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   26-57             #domain EGF homology #label EGF
SUMMARY          #length 861  #checksum 7963

  Query Match           24.5%;  Score 198;  DB 2;  Length 861;
  Best Local Similarity  40.0%;  Pred. No. 4.94e-24;
  Matches   32;  Conservative   19;  Mismatches  22;  Indels   7;  Gaps   4;

Db    158 NECASNPCQNGANCTDCVDSYTCTCPVGFNGIHCENNTPDCT-ESS-C--F---NGGTCV 210
          ::| :: |::|| | | |::||| || ||:|: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    211 DGINSFTCLCPPGFTGSYCQ 230
                || |||||:|: |:
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT   13
ENTRY            A46019     #type complete
TITLE            gene Notch-1 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                   14-Aug-1998
ACCESSIONS       A46019
REFERENCE        A46019
   #authors      del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins,
                   N.A.; Copeland, N.G.; Gridley, T.
   #journal      Genomics (1993) 15:259-264
   #title        Cloning, analysis, and chromosomal localization of Notch-1, a
                   mouse homolog of Drosophila Notch.
   #cross-references MUID:93194170
   #accession    A46019
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues     1-2531 ##label DEL
      ##cross-references GB:Z11886; GB:S47228; NID:g288502; PID:g288503
      ##note         sequence extracted from NCBI backbone (NCBIP:127318)
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
```

```
FEATURE
   757-788          #domain EGF homology #label EGF\
   1917-1948        #domain ankyrin repeat homology #label AN1\
   1949-1981        #domain ankyrin repeat homology #label AN2\
   1983-2015        #domain ankyrin repeat homology #label AN3\
   2016-2048        #domain ankyrin repeat homology #label AN4\
   2049-2081        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2531  #molecular-weight 271312  #checksum 6611

  Query Match           24.5%;  Score 198;  DB 2;  Length 2531;
  Best Local Similarity  40.0%;  Pred. No. 4.94e-24;
  Matches    32;  Conservative   19;  Mismatches  22;  Indels   7;  Gaps   4;

Db    947 NECASNPCQNGANCTDCVDSYTCTCPVGFNGIHCENNTPDCT-ESS-C--F---NGGTCV 999
          ::| :: |::|| | | |::||| || ||:|: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINSFTCLCPPGFTGSYCQ 1019
             || |||||:|: |:
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT   14
ENTRY            S18188     #type complete
TITLE            notch protein homolog - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
                   12-Feb-1999
ACCESSIONS       S18188
REFERENCE        S18188
   #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal      Development (1991) 113:199-205
   #title        A homolog of Drosophila Notch expressed during mammalian
                   development.
   #cross-references MUID:92111383
   #accession    S18188
      ##molecule_type mRNA
      ##residues      1-2531 ##label WEI
      ##cross-references EMBL:X57405; NID:g57634; PID:g57635
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1917-1949        #domain ankyrin repeat homology #label AN1\
   1950-1982        #domain ankyrin repeat homology #label AN2\
   1984-2016        #domain ankyrin repeat homology #label AN3\
   2017-2049        #domain ankyrin repeat homology #label AN4\
   2050-2082        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2531  #molecular-weight 270907  #checksum 2705

  Query Match           24.5%;  Score 198;  DB 2;  Length 2531;
  Best Local Similarity  40.0%;  Pred. No. 4.94e-24;
  Matches    32;  Conservative   18;  Mismatches  23;  Indels   7;  Gaps   4;

Db    947 NECATNPCQNGANCTDCVDSYTCTCPTGFNGIHCENNTPDCT-ESS-C--F---NGGTCV 999
          ::|  : |::|| | | |::||| || ||:|: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINSFTCLCPPGFTGSYCQ 1019
             || |||||:|: |:
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT   15
ENTRY            B49175     #type fragment
TITLE            Motch A protein - mouse (fragment)
ALTERNATE_NAMES  Notch homolog
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                   14-Aug-1998
ACCESSIONS       B49175; PH1569; S32109
REFERENCE        A49175
   #authors      Lardelli, M.; Lendahl, U.
   #journal      Exp. Cell Res. (1993) 204:364-372
   #title        Motch A and Motch B--two mouse Notch homologues coexpressed
                   in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession    B49175
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-387 ##label LAR
      ##cross-references EMBL:X68278; NID:g287987; PID:g287988
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:126159)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermal and neural fate for cells in the
             early embryo.
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
FEATURE
   27-58            #domain EGF homology #label EGF
SUMMARY          #length 387  #checksum 5404

  Query Match           24.4%;  Score 197;  DB 2;  Length 387;
  Best Local Similarity  35.8%;  Pred. No. 8.20e-24;
  Matches    29;  Conservative   15;  Mismatches  36;  Indels   1;  Gaps   1;

Db     25 NECLSQPCQNGGTCIDLTNSYKCSCPRGTQGVHCEINVDDCHPPLDPASRSPKCFNNGTC 84
          ::|::: |::|: |:|  |:| | ||:|  |: ||          |  :   | | : |
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQY-ECQNGAQC 61

Db     85 VDQVGGYTCTCPPGFVGERCE 105
          :      || ||||| | |||
Qy     62 IVVQQEPTCRCPPGFAGPRCE 82


Search completed: Fri May 28 09:35:38 1999
Job time : 22 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

IPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:35:57 1999; MasPar time 5.04 Seconds
583.869 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-13
Description: (1-104) from US09191647.pep
Perfect Score: 807
Sequence: 1 NNDDCVGHKCRHGAQCVDEV.....ITVNFVGKDSYVELASAKVR 104

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 37.259; Variance 53.097; scale 0.702

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	343	42.5	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	7.90e-66
2	215	26.6	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	6.84e-32
3	214	26.5	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	1.23e-31
4	209	25.9	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	2.28e-30
5	205	25.4	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	2.33e-29
6	205	25.4	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	2.33e-29
7	202	25.0	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	1.32e-28
8	198	24.5	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	1.33e-27
9	198	24.5	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	1.33e-27
10	195	24.2	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	7.46e-27
11	193	23.9	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	2.35e-26
12	192	23.8	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	4.16e-26
13	190	23.5	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	1.30e-25
14	188	23.3	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	4.08e-25
15	188	23.3	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	4.08e-25
16	185	22.9	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	2.24e-24
17	185	22.9	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	2.24e-24
18	184	22.8	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	3.95e-24
19	175	21.7	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	6.27e-22
20	171	21.2	492	1	FA10_BOVIN	COAGULATION FACTOR X P	5.84e-21
21	171	21.2	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	5.84e-21
22	171	21.2	4543	1	LRP1_CHICK	LOW-DENSITY LIPOPROTEI	5.84e-21
23	165	20.4	475	1	FA10_CHICK	COAGULATION FACTOR X P	1.62e-19
24	164	20.3	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	2.80e-19
25	163	20.2	3051	1	YNX3_CAEEL	HYPOTHETICAL PROTEIN T	4.86e-19
26	162	20.1	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	8.41e-19
27	154	19.1	444	1	FA7_RABIT	COAGULATION FACTOR VII	6.55e-17
28	152	18.8	461	1	FA9_HUMAN	COAGULATION FACTOR IX	1.93e-16
29	152	18.8	488	1	FA10_HUMAN	COAGULATION FACTOR X P	1.93e-16
30	150	18.6	407	1	FA7_BOVIN	COAGULATION FACTOR VII	5.64e-16
31	150	18.6	3562	1	PGCV_CHICK	VERSICAN CORE PROTEIN	5.64e-16
32	148	18.3	409	1	MFGM_PIG	MILK FAT GLOBULE-EGF F	1.64e-15
33	147	18.2	427	1	MFGM_BOVIN	MILK FAT GLOBULE-EGF F	2.80e-15
34	146	18.1	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	4.77e-15
35	144	17.8	643	1	UROM_BOVIN	UROMODULIN PRECURSOR (	1.38e-14
36	144	17.8	1257	1	PGCN_RAT	NEUROCAN CORE PROTEIN	1.38e-14
37	144	17.8	1268	1	PGCN_MOUSE	NEUROCAN CORE PROTEIN	1.38e-14
38	143	17.7	4544	1	LRP1_HUMAN	LOW-DENSITY LIPOPROTEI	2.34e-14
39	142	17.6	466	1	FA7_HUMAN	COAGULATION FACTOR VII	3.96e-14
40	142	17.6	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	3.96e-14
41	142	17.6	3358	1	PGCV_MOUSE	VERSICAN CORE PROTEIN	3.96e-14
42	142	17.6	3396	1	PGCV_HUMAN	VERSICAN CORE PROTEIN	3.96e-14
43	141	17.5	416	1	FA9_BOVIN	COAGULATION FACTOR IX	6.71e-14
44	141	17.5	610	1	LEM2_HUMAN	E-SELECTIN PRECURSOR (	6.71e-14
45	141	17.5	2871	1	FBN1_BOVIN	FIBRILLIN 1 PRECURSOR	6.71e-14

ALIGNMENTS

RESULT 1
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91099665.
RA ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL GENES DEV. 4:2169-2187(1990).
CC -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X53959; G8615; -.
DR PIR; A36665; A36665.
DR FLYBASE; FBgn0003425; sli.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01185; CTCK_1; 1.

```
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1    36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   DOMAIN      70   104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105   230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231   294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295   326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327   452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453   518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519   550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551   653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654   714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715   746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747   848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849   910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105   115       LRR 1-1.
FT   REPEAT     116   139       LRR 1-2.
FT   REPEAT     140   163       LRR 1-3.
FT   REPEAT     164   187       LRR 1-4.
FT   REPEAT     188   211       LRR 1-5.
FT   REPEAT     212   230       LRR 1-6.
FT   REPEAT     327   337       LRR 2-1.
FT   REPEAT     338   361       LRR 2-2.
FT   REPEAT     362   385       LRR 2-3.
FT   REPEAT     386   409       LRR 2-4.
FT   REPEAT     410   433       LRR 2-5.
FT   REPEAT     434   452       LRR 2-6.
FT   REPEAT     551   562       LRR 3-1.
FT   REPEAT     563   586       LRR 3-2.
FT   REPEAT     587   610       LRR 3-3.
FT   REPEAT     611   634       LRR 3-4.
FT   REPEAT     635   653       LRR 3-5.
FT   REPEAT     747   757       LRR 4-1.
FT   REPEAT     758   781       LRR 4-2.
FT   REPEAT     782   805       LRR 4-3.
FT   REPEAT     806   829       LRR 4-4.
FT   REPEAT     830   848       LRR 4-5.
FT   DOMAIN     907   944       EGF-LIKE 1.
FT   DOMAIN     946   983       EGF-LIKE 2.
FT   DOMAIN     985  1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062       EGF-LIKE 4.
FT   DOMAIN    1064  1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149       EGF-LIKE 6.
FT   DOMAIN    1353  1392       EGF-LIKE 7.
FT   DOMAIN    1409  1480       CTCK.
FT   VARSPLIC  1394  1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111   111       POTENTIAL.
FT   CARBOHYD   207   207       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
FT   CARBOHYD   435   435       POTENTIAL.
FT   CARBOHYD   783   783       POTENTIAL.
FT   CARBOHYD   788   788       POTENTIAL.
FT   CARBOHYD   958   958       POTENTIAL.
FT   CARBOHYD   998   998       POTENTIAL.
FT   CARBOHYD  1060  1060       POTENTIAL.
FT   CARBOHYD  1159  1159       POTENTIAL.
FT   CARBOHYD  1175  1175       POTENTIAL.
FT   CARBOHYD  1243  1243       POTENTIAL.
FT   CARBOHYD  1292  1292       POTENTIAL.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   932       BY SIMILARITY.
FT   DISULFID   934   943       BY SIMILARITY.
FT   DISULFID   950   961       BY SIMILARITY.
FT   DISULFID   955   971       BY SIMILARITY.
FT   DISULFID   973   982       BY SIMILARITY.
FT   DISULFID   989  1001       BY SIMILARITY.
FT   DISULFID   995  1010       BY SIMILARITY.
FT   DISULFID  1012  1021       BY SIMILARITY.
FT   DISULFID  1028  1041       BY SIMILARITY.
FT   DISULFID  1035  1050       BY SIMILARITY.
FT   DISULFID  1052  1061       BY SIMILARITY.
FT   DISULFID  1068  1079       BY SIMILARITY.
FT   DISULFID  1073  1088       BY SIMILARITY.
FT   DISULFID  1090  1099       BY SIMILARITY.
FT   DISULFID  1115  1125       BY SIMILARITY.
FT   DISULFID  1120  1137       BY SIMILARITY.
FT   DISULFID  1139  1148       BY SIMILARITY.
FT   DISULFID  1357  1368       BY SIMILARITY.
FT   DISULFID  1362  1380       BY SIMILARITY.
FT   DISULFID  1382  1391       BY SIMILARITY.
FT   DISULFID  1409  1443       BY SIMILARITY.
FT   DISULFID  1423  1457       BY SIMILARITY.
FT   DISULFID  1434  1473       BY SIMILARITY.
FT   DISULFID  1438  1475       BY SIMILARITY.
FT   DISULFID  1442  1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          42.5%;  Score 343;  DB 1;  Length 1480;
  Best Local Similarity  42.2%;  Pred. No. 7.90e-66;
  Matches   46;  Conservative   26;  Mismatches  31;  Indels   6;  Gaps   5;

Db   1064 NIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHG 1123
            | |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db   1124 V-CFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTR 1171
             |:    | :  ||| ||::| :|| | :::|| ::|:||| : : |
Qy     59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    2
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19  >2444       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737   1757       POTENTIAL.
FT   DOMAIN    1758  >2444       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20.
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    868       EGF-LIKE 22.
FT   DOMAIN     870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     908    944       EGF-LIKE 24.
FT   DOMAIN     946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 26.
FT   DOMAIN    1022   1058       EGF-LIKE 27.
FT   DOMAIN    1060   1096       EGF-LIKE 28.
FT   DOMAIN    1098   1144       EGF-LIKE 29.
FT   DOMAIN    1146   1182       EGF-LIKE 30.
FT   DOMAIN    1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1268   1306       EGF-LIKE 33.
FT   DOMAIN    1308   1347       EGF-LIKE 34.
FT   DOMAIN    1349   1385       EGF-LIKE 35.
FT   DOMAIN    1388   1427       EGF-LIKE 36.
FT   DOMAIN    1446   1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446   1481       LIN/NOTCH 1.
FT   REPEAT    1482   1523       LIN/NOTCH 2.
FT   REPEAT    1524   1563       LIN/NOTCH 3.
FT   DOMAIN    1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT    1876   1921       ANK MOTIF 1.
FT   REPEAT    1923   1954       ANK MOTIF 2.
FT   REPEAT    1956   1987       ANK MOTIF 3.
FT   REPEAT    1990   2021       ANK MOTIF 4.
FT   REPEAT    2023   2054       ANK MOTIF 5.
FT   REPEAT    2056   2087       ANK MOTIF 6.
FT   DOMAIN    1576   1579       POLY-VAL.
FT   DOMAIN    1662   1665       POLY-ARG.
FT   DOMAIN    1729   1732       POLY-PRO.
FT   DOMAIN    1741   1744       POLY-ALA.
FT   DOMAIN    1902   1905       POLY-GLU.
FT   DOMAIN    2260   2263       POLY-GLY.
FT   DOMAIN    2404   2407       POLY-GLN.
FT   DOMAIN    2411   2418       POLY-PRO.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
FT   DISULFID   339    350       BY SIMILARITY.
FT   DISULFID   344    359       BY SIMILARITY.
FT   DISULFID   361    370       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    398       BY SIMILARITY.
FT   DISULFID   400    409       BY SIMILARITY.
FT   DISULFID   416    429       BY SIMILARITY.
FT   DISULFID   423    438       BY SIMILARITY.
FT   DISULFID   440    449       BY SIMILARITY.
FT   DISULFID   456    467       BY SIMILARITY.
FT   DISULFID   461    476       BY SIMILARITY.
FT   DISULFID   478    487       BY SIMILARITY.
FT   DISULFID   494    505       BY SIMILARITY.
FT   DISULFID   499    514       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   532    543       BY SIMILARITY.
FT   DISULFID   537    552       BY SIMILARITY.
FT   DISULFID   554    563       BY SIMILARITY.
FT   DISULFID   570    580       BY SIMILARITY.
FT   DISULFID   575    589       BY SIMILARITY.
FT   DISULFID   591    600       BY SIMILARITY.
FT   DISULFID   607    618       BY SIMILARITY.
FT   DISULFID   612    627       BY SIMILARITY.
FT   DISULFID   629    638       BY SIMILARITY.
FT   DISULFID   645    655       BY SIMILARITY.
FT   DISULFID   650    664       BY SIMILARITY.
FT   DISULFID   666    675       BY SIMILARITY.
FT   DISULFID   682    693       BY SIMILARITY.
FT   DISULFID   687    702       BY SIMILARITY.
FT   DISULFID   704    713       BY SIMILARITY.
FT   DISULFID   720    730       BY SIMILARITY.
FT   DISULFID   725    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    855       BY SIMILARITY.
FT   DISULFID   857    867       BY SIMILARITY.
FT   DISULFID   874    885       BY SIMILARITY.
FT   DISULFID   879    894       BY SIMILARITY.
FT   DISULFID   896    905       BY SIMILARITY.
FT   DISULFID   912    923       BY SIMILARITY.
FT   DISULFID   917    932       BY SIMILARITY.
```

```
FT   DISULFID    934    943       BY SIMILARITY.
FT   DISULFID    988    999       BY SIMILARITY.
FT   DISULFID    993   1008       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1026   1037       BY SIMILARITY.
FT   DISULFID   1031   1046       BY SIMILARITY.
FT   DISULFID   1048   1057       BY SIMILARITY.
FT   DISULFID   1064   1075       BY SIMILARITY.
FT   DISULFID   1069   1084       BY SIMILARITY.
FT   DISULFID   1086   1095       BY SIMILARITY.
FT   DISULFID   1102   1123       BY SIMILARITY.
FT   DISULFID   1117   1132       BY SIMILARITY.
FT   DISULFID   1134   1143       BY SIMILARITY.
FT   DISULFID   1150   1161       BY SIMILARITY.
FT   DISULFID   1155   1170       BY SIMILARITY.
FT   DISULFID   1172   1181       BY SIMILARITY.
FT   DISULFID   1188   1199       BY SIMILARITY.
FT   DISULFID   1193   1208       BY SIMILARITY.
...
Note: remainder of annotations omitted.

[redacted]uery Match          26.6%;  Score 215;  DB 1;  Length 2444;
Best Local Similarity  41.0%;  Pred. No. 6.84e-32;
  Matches   34;  Conservative   18;  Mismatches  24;  Indels   7;  Gaps   4;

Db    948 NECASDPCRNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCT-ESS-C--F---NGGTCV 1000
          ::| :  ||:|| | | |::||| || ||||: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1001 DGINSFTCLCPPGFTGSYCQHVV 1023
               || |||||:|: |: ::
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85


RESULT    3
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1064       FIBROPELLIN I.
FT   DOMAIN      20     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480    516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     518    554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     556    592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     594    630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632    668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     670    706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     708    744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     746    782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     784    820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     822    858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     860    896       EGF-LIKE 20.
FT   DOMAIN     898    934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     936   1064       AVIDIN-LIKE.
FT   DISULFID    23     34       BY SIMILARITY.
```

```
FT   DISULFID    28     43       BY SIMILARITY.
FT   DISULFID    45     54       BY SIMILARITY.
FT   DISULFID   180    191       BY SIMILARITY.
FT   DISULFID   185    200       BY SIMILARITY.
FT   DISULFID   202    211       BY SIMILARITY.
FT   DISULFID   218    229       BY SIMILARITY.
FT   DISULFID   223    238       BY SIMILARITY.
FT   DISULFID   240    249       BY SIMILARITY.
FT   DISULFID   256    267       BY SIMILARITY.
FT   DISULFID   261    276       BY SIMILARITY.
FT   DISULFID   278    287       BY SIMILARITY.
FT   DISULFID   294    305       BY SIMILARITY.
FT   DISULFID   299    314       BY SIMILARITY.
FT   DISULFID   316    325       BY SIMILARITY.
FT   DISULFID   332    343       BY SIMILARITY.
FT   DISULFID   337    352       BY SIMILARITY.
FT   DISULFID   354    363       BY SIMILARITY.
FT   DISULFID   370    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   430    439       BY SIMILARITY.
FT   DISULFID   446    457       BY SIMILARITY.
FT   DISULFID   451    466       BY SIMILARITY.
FT   DISULFID   468    477       BY SIMILARITY.
FT   DISULFID   484    495       BY SIMILARITY.
FT   DISULFID   489    504       BY SIMILARITY.
FT   DISULFID   506    515       BY SIMILARITY.
FT   DISULFID   522    533       BY SIMILARITY.
FT   DISULFID   527    542       BY SIMILARITY.
FT   DISULFID   544    553       BY SIMILARITY.
FT   DISULFID   560    571       BY SIMILARITY.
FT   DISULFID   565    580       BY SIMILARITY.
FT   DISULFID   582    591       BY SIMILARITY.
FT   DISULFID   598    609       BY SIMILARITY.
FT   DISULFID   603    618       BY SIMILARITY.
FT   DISULFID   620    629       BY SIMILARITY.
FT   DISULFID   636    647       BY SIMILARITY.
FT   DISULFID   641    656       BY SIMILARITY.
FT   DISULFID   658    667       BY SIMILARITY.
FT   DISULFID   674    685       BY SIMILARITY.
FT   DISULFID   679    694       BY SIMILARITY.
FT   DISULFID   696    705       BY SIMILARITY.
FT   DISULFID   712    723       BY SIMILARITY.
FT   DISULFID   717    732       BY SIMILARITY.
FT   DISULFID   734    743       BY SIMILARITY.
FT   DISULFID   750    761       BY SIMILARITY.
FT   DISULFID   755    770       BY SIMILARITY.
FT   DISULFID   772    781       BY SIMILARITY.
FT   DISULFID   788    799       BY SIMILARITY.
FT   DISULFID   793    808       BY SIMILARITY.
FT   DISULFID   810    819       BY SIMILARITY.
FT   DISULFID   826    837       BY SIMILARITY.
FT   DISULFID   831    846       BY SIMILARITY.
FT   DISULFID   848    857       BY SIMILARITY.
FT   DISULFID   864    875       BY SIMILARITY.
FT   DISULFID   869    884       BY SIMILARITY.
FT   DISULFID   886    895       BY SIMILARITY.
FT   DISULFID   902    913       BY SIMILARITY.
FT   DISULFID   907    922       BY SIMILARITY.
FT   DISULFID   924    933       BY SIMILARITY.
FT   VARSPLIC   477    780       MISSING (IN FORM IB).
FT   CARBOHYD    30     30       POTENTIAL.
FT   CARBOHYD   136    136       POTENTIAL.
FT   CARBOHYD   851    851       POTENTIAL.
FT   CONFLICT   279    279       L -> S (IN REF. 2).
SQ   SEQUENCE  1064 AA;  112072 MW;  FBD10D48 CRC32;

  Query Match          26.5%;  Score 214;  DB 1;  Length 1064;
  Best Local Similarity  42.7%;  Pred. No. 1.23e-31;
  Matches   35;  Conservative   13;  Mismatches  27;  Indels   7;  Gaps   2;

Db    822 NIDECASDPCLNGGICVDGVNGFVCQCPPNYSGTYCE-----ISLDA--CRSMPCQNGAT 874
          | |:| :  | :|: ||| |||: | ||  :|| :||     : |:: |    |||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    875 CVNVGADYVCECVPGYAGQNCE 896
          |: |  :  | | ||:||  ||
Qy     61 CIVVQQEPTCRCPPGFAGPRCE 82


RESULT    4
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     44       POTENTIAL.
FT   CHAIN       45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746   1766       POTENTIAL.
FT   DOMAIN    1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58     95       EGF-LIKE 1.
FT   DOMAIN      96    136       EGF-LIKE 2.
FT   DOMAIN     139    176       EGF-LIKE 3.
FT   DOMAIN     177    215       EGF-LIKE 4.
FT   DOMAIN     217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255    291       EGF-LIKE 6.
FT   DOMAIN     293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     409    447       EGF-LIKE 10.
FT   DOMAIN     449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     867    905       EGF-LIKE 22.
FT   DOMAIN     907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 25.
FT   DOMAIN    1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1060   1096       EGF-LIKE 27.
FT   DOMAIN    1098   1134       EGF-LIKE 28.
FT   DOMAIN    1136   1181       EGF-LIKE 29.
FT   DOMAIN    1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1297   1335       EGF-LIKE 33.
FT   DOMAIN    1337   1373       EGF-LIKE 34.
FT   DOMAIN    1375   1412       EGF-LIKE 35.
FT   DOMAIN    1415   1451       EGF-LIKE 36.
FT   DOMAIN    1475   1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1475   1513       LIN/NOTCH 1.
FT   REPEAT    1514   1553       LIN/NOTCH 2.
FT   REPEAT    1554   1593       LIN/NOTCH 3.
FT   DOMAIN    1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN    2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     83       BY SIMILARITY.
FT   DISULFID    85     94       BY SIMILARITY.
FT   DISULFID   100    111       BY SIMILARITY.
FT   DISULFID   105    124       BY SIMILARITY.
FT   DISULFID   126    135       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   181    192       BY SIMILARITY.
FT   DISULFID   186    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    241       BY SIMILARITY.
FT   DISULFID   243    252       BY SIMILARITY.
FT   DISULFID   259    270       BY SIMILARITY.
FT   DISULFID   264    279       BY SIMILARITY.
FT   DISULFID   281    290       BY SIMILARITY.
FT   DISULFID   297    308       BY SIMILARITY.
FT   DISULFID   302    317       BY SIMILARITY.
FT   DISULFID   319    328       BY SIMILARITY.
FT   DISULFID   335    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    396       BY SIMILARITY.
FT   DISULFID   398    407       BY SIMILARITY.
FT   DISULFID   413    424       BY SIMILARITY.
FT   DISULFID   418    435       BY SIMILARITY.
FT   DISULFID   437    446       BY SIMILARITY.
FT   DISULFID   453    465       BY SIMILARITY.
FT   DISULFID   459    474       BY SIMILARITY.
FT   DISULFID   476    485       BY SIMILARITY.
FT   DISULFID   492    503       BY SIMILARITY.
FT   DISULFID   497    512       BY SIMILARITY.
FT   DISULFID   514    523       BY SIMILARITY.
FT   DISULFID   530    541       BY SIMILARITY.
FT   DISULFID   535    550       BY SIMILARITY.
FT   DISULFID   552    561       BY SIMILARITY.
FT   DISULFID   568    579       BY SIMILARITY.
FT   DISULFID   573    588       BY SIMILARITY.
FT   DISULFID   590    599       BY SIMILARITY.
FT   DISULFID   606    616       BY SIMILARITY.
FT   DISULFID   611    625       BY SIMILARITY.
FT   DISULFID   627    636       BY SIMILARITY.
FT   DISULFID   643    654       BY SIMILARITY.
FT   DISULFID   648    663       BY SIMILARITY.
FT   DISULFID   665    674       BY SIMILARITY.
FT   DISULFID   681    692       BY SIMILARITY.
FT   DISULFID   686    701       BY SIMILARITY.
FT   DISULFID   703    712       BY SIMILARITY.
FT   DISULFID   719    730       BY SIMILARITY.
FT   DISULFID   724    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    853       BY SIMILARITY.
FT   DISULFID   855    864       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          25.9%;  Score 209;  DB 1;  Length 2703;
  Best Local Similarity  40.7%;  Pred. No. 2.28e-30;
  Matches   33;  Conservative   14;  Mismatches  30;  Indels   4;  Gaps   2;

Db.   293 NYDDCLGHLCQNGGTCIDGISDYTCRCPPNFTGRFCQDD---VD-ECAQRDHPVCQNGAT 348
          | |||:|| |::|: |:| :: ||| ||  |:| ||:     | : :  |:  |||||
```

```
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    349 CTNTHGSYSCICVNGWAGLDC 369
          |   :   :| |  | ||  |
Qy     61 CIVVQQEPTCRCPPGFAGPRC 81


RESULT    5
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1     90
FT   CHAIN       91   2139       CRUMBS PROTEIN.
FT   DOMAIN      91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085   2111       POTENTIAL.
FT   DOMAIN    2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267    303       EGF-LIKE 1.
FT   DOMAIN     306    343       EGF-LIKE 2.
FT   DOMAIN     348    386       EGF-LIKE 3.
FT   DOMAIN     388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427    463       EGF-LIKE 5.
FT   DOMAIN     464    500       EGF-LIKE 6.
FT   DOMAIN     501    532       EGF-LIKE 7.
FT   DOMAIN     545    581       EGF-LIKE 8.
FT   DOMAIN     582    611       EGF-LIKE 9.
FT   DOMAIN     609    646       EGF-LIKE 10.
FT   DOMAIN     648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840    902       EGF-LIKE 16.
FT   DOMAIN     904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980   1021       EGF-LIKE 19.
FT   DOMAIN    1207   1243       EGF-LIKE 20.
FT   DOMAIN    1481   1517       EGF-LIKE 21.
FT   DOMAIN    1759   1795       EGF-LIKE 22.
FT   DOMAIN    1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874   1915       EGF-LIKE 25.
FT   DOMAIN    1915   1951       EGF-LIKE 26.
FT   DOMAIN    1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030   2070       EGF-LIKE 29.
FT   DISULFID   271    282       BY SIMILARITY.
FT   DISULFID   276    291       BY SIMILARITY.
FT   DISULFID   293    302       BY SIMILARITY.
FT   DISULFID   310    321       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   333    342       BY SIMILARITY.
FT   DISULFID   352    363       BY SIMILARITY.
FT   DISULFID   357    374       BY SIMILARITY.
FT   DISULFID   376    385       BY SIMILARITY.
FT   DISULFID   392    403       BY SIMILARITY.
FT   DISULFID   397    412       BY SIMILARITY.
FT   DISULFID   414    424       BY SIMILARITY.
FT   DISULFID   431    442       BY SIMILARITY.
FT   DISULFID   436    451       BY SIMILARITY.
FT   DISULFID   453    462       BY SIMILARITY.
FT   DISULFID   468    479       BY SIMILARITY.
FT   DISULFID   473    488       BY SIMILARITY.
FT   DISULFID   490    499       BY SIMILARITY.
FT   DISULFID   505    515       BY SIMILARITY.
FT   DISULFID   509    520       BY SIMILARITY.
FT   DISULFID   522    531       BY SIMILARITY.
FT   DISULFID   549    562       BY SIMILARITY.
FT   DISULFID   556    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   586    597       BY SIMILARITY.
FT   DISULFID   591    602       BY SIMILARITY.
FT   DISULFID   604    610       BY SIMILARITY.
FT   DISULFID   613    624       BY SIMILARITY.
FT   DISULFID   618    634       BY SIMILARITY.
FT   DISULFID   636    645       BY SIMILARITY.
FT   DISULFID   652    664       BY SIMILARITY.
FT   DISULFID   659    673       BY SIMILARITY.
FT   DISULFID   675    684       BY SIMILARITY.
FT   DISULFID   691    702       BY SIMILARITY.
FT   DISULFID   696    711       BY SIMILARITY.
FT   DISULFID   713    722       BY SIMILARITY.
FT   DISULFID   729    740       BY SIMILARITY.
FT   DISULFID   734    749       BY SIMILARITY.
FT   DISULFID   751    760       BY SIMILARITY.
FT   DISULFID   767    778       BY SIMILARITY.
FT   DISULFID   772    787       BY SIMILARITY.
FT   DISULFID   789    799       BY SIMILARITY.
FT   DISULFID   806    817       BY SIMILARITY.
FT   DISULFID   811    826       BY SIMILARITY.
FT   DISULFID   828    837       BY SIMILARITY.
FT   DISULFID   844    855       BY SIMILARITY.
```

```
FT   DISULFID    849    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID    908    919       BY SIMILARITY.
FT   DISULFID    913    928       BY SIMILARITY.
FT   DISULFID    930    939       BY SIMILARITY.
FT   DISULFID    946    957       BY SIMILARITY.
FT   DISULFID    952    966       BY SIMILARITY.
FT   DISULFID    968    977       BY SIMILARITY.
FT   DISULFID    984    995       BY SIMILARITY.
FT   DISULFID    989   1009       BY SIMILARITY.
FT   DISULFID   1011   1020       BY SIMILARITY.
FT   DISULFID   1211   1222       BY SIMILARITY.
FT   DISULFID   1216   1231       BY SIMILARITY.
FT   DISULFID   1233   1242       BY SIMILARITY.
FT   DISULFID   1485   1496       BY SIMILARITY.
FT   DISULFID   1490   1505       BY SIMILARITY.
FT   DISULFID   1507   1516       BY SIMILARITY.
FT   DISULFID   1763   1774       BY SIMILARITY.
FT   DISULFID   1768   1783       BY SIMILARITY.
FT   DISULFID   1785   1794       BY SIMILARITY.
FT   DISULFID   1801   1812       BY SIMILARITY.
FT   DISULFID   1806   1821       BY SIMILARITY.
FT   DISULFID   1823   1832       BY SIMILARITY.
FT   DISULFID   1839   1850       BY SIMILARITY.
FT   DISULFID   1844   1859       BY SIMILARITY.
FT   DISULFID   1861   1870       BY SIMILARITY.
FT   DISULFID   1878   1889       BY SIMILARITY.
FT   DISULFID   1883   1903       BY SIMILARITY.
FT   DISULFID   1905   1914       BY SIMILARITY.
FT   DISULFID   1919   1930       BY SIMILARITY.
FT   DISULFID   1924   1939       BY SIMILARITY.
FT   DISULFID   1941   1950       BY SIMILARITY.
FT   DISULFID   1957   1968       BY SIMILARITY.
FT   DISULFID   1962   1977       BY SIMILARITY.
FT   DISULFID   1979   1988       BY SIMILARITY.
FT   DISULFID   1995   2008       BY SIMILARITY.
FT   DISULFID   2002   2017       BY SIMILARITY.
FT   DISULFID   2019   2028       BY SIMILARITY.
FT   CARBOHYD     37     37       POTENTIAL.
FT   CARBOHYD     96     96       POTENTIAL.
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    238    238       POTENTIAL.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    336    336       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CARBOHYD    550    550       POTENTIAL.
FT   CARBOHYD    565    565       POTENTIAL.
FT   CARBOHYD    736    736       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    860    860       POTENTIAL.
FT   CARBOHYD    884    884       POTENTIAL.
FT   CARBOHYD    976    976       POTENTIAL.
FT   CARBOHYD   1102   1102       POTENTIAL.
FT   CARBOHYD   1114   1114       POTENTIAL.
FT   CARBOHYD   1138   1138       POTENTIAL.
FT   CARBOHYD   1192   1192       POTENTIAL.
FT   CARBOHYD   1245   1245       POTENTIAL.
FT   CARBOHYD   1255   1255       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   CARBOHYD   1363   1363       POTENTIAL.
FT   CARBOHYD   1441   1441       POTENTIAL.
FT   CARBOHYD   1454   1454       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          25.4%;  Score 205;  DB 1;  Length 2139;
  Best Local Similarity  34.9%;  Pred. No. 2.33e-29;
  Matches   29;  Conservative   14;  Mismatches  37;  Indels   3;  Gaps   3;

Db   1835 NIDECADQPCHNGGNCTDLIASYVCDCPEDYMGPQCDVLKQMTC-ENEPCRNGSTCQNGF 1893
          | |:|  : |::|: | | : :| | ||: : |  |:    |   :  || :   ||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPC-DQYECQNGA 59

Db   1894 N-ASTGNNFTCTCVPGFEGPLCD 1915
                : || | ||| || |:
Qy     60 QCIVVQQEPTCRCPPGFAGPRCE 82


RESULT    6
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20   1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729   1750       POTENTIAL.
FT   DOMAIN    1751   2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     57       EGF-LIKE 1.
FT   DOMAIN      58     99       EGF-LIKE 2.
FT   DOMAIN     102    140       EGF-LIKE 3.
FT   DOMAIN     141    177       EGF-LIKE 4.
FT   DOMAIN     179    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527    563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565    600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      602    638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      640    675       EGF-LIKE 17.
FT   DOMAIN      677    713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      715    750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      752    788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      790    826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      828    866       EGF-LIKE 22.
FT   DOMAIN      868    904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      906    942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      944    980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      982   1018       EGF-LIKE 26.
FT   DOMAIN     1020   1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1058   1094       EGF-LIKE 28.
FT   DOMAIN     1096   1142       EGF-LIKE 29.
FT   DOMAIN     1144   1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1182   1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1220   1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1266   1304       EGF-LIKE 33.
FT   DOMAIN     1306   1346       EGF-LIKE 34.
FT   DOMAIN     1347   1383       EGF-LIKE 35.
FT   DOMAIN     1386   1424       EGF-LIKE 36.
FT   DOMAIN     1441   1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1441   1478       LIN/NOTCH 1.
FT   REPEAT     1479   1520       LIN/NOTCH 2.
FT   REPEAT     1521   1560       LIN/NOTCH 3.
FT   DOMAIN     1871   2083       6 X ANK MOTIF REPEATS.
FT   DISULFID     22     35       BY SIMILARITY.
FT   DISULFID     29     45       BY SIMILARITY.
FT   DISULFID     47     56       BY SIMILARITY.
FT   DISULFID     62     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    128       BY SIMILARITY.
FT   DISULFID    130    139       BY SIMILARITY.
FT   DISULFID    145    156       BY SIMILARITY.
FT   DISULFID    150    165       BY SIMILARITY.
FT   DISULFID    167    176       BY SIMILARITY.
FT   DISULFID    183    194       BY SIMILARITY.
FT   DISULFID    188    203       BY SIMILARITY.
FT   DISULFID    205    214       BY SIMILARITY.
FT   DISULFID    221    232       BY SIMILARITY.
FT   DISULFID    226    242       BY SIMILARITY.
FT   DISULFID    244    253       BY SIMILARITY.
FT   DISULFID    260    271       BY SIMILARITY.
FT   DISULFID    265    280       BY SIMILARITY.
FT   DISULFID    282    291       BY SIMILARITY.
FT   DISULFID    298    311       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    322    331       BY SIMILARITY.
FT   DISULFID    338    349       BY SIMILARITY.
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    375    386       BY SIMILARITY.
FT   DISULFID    380    397       BY SIMILARITY.
FT   DISULFID    399    408       BY SIMILARITY.
FT   DISULFID    415    428       BY SIMILARITY.
FT   DISULFID    422    437       BY SIMILARITY.
FT   DISULFID    439    448       BY SIMILARITY.
FT   DISULFID    455    466       BY SIMILARITY.
FT   DISULFID    460    475       BY SIMILARITY.
FT   DISULFID    477    486       BY SIMILARITY.
FT   DISULFID    493    504       BY SIMILARITY.
FT   DISULFID    498    513       BY SIMILARITY.
FT   DISULFID    515    524       BY SIMILARITY.
FT   DISULFID    531    542       BY SIMILARITY.
FT   DISULFID    536    551       BY SIMILARITY.
FT   DISULFID    553    562       BY SIMILARITY.
FT   DISULFID    569    579       BY SIMILARITY.
FT   DISULFID    574    588       BY SIMILARITY.
FT   DISULFID    590    599       BY SIMILARITY.
FT   DISULFID    606    617       BY SIMILARITY.
FT   DISULFID    611    626       BY SIMILARITY.
FT   DISULFID    628    637       BY SIMILARITY.
FT   DISULFID    644    654       BY SIMILARITY.
FT   DISULFID    649    663       BY SIMILARITY.
FT   DISULFID    665    674       BY SIMILARITY.
FT   DISULFID    681    692       BY SIMILARITY.
FT   DISULFID    686    701       BY SIMILARITY.
FT   DISULFID    703    712       BY SIMILARITY.
FT   DISULFID    719    729       BY SIMILARITY.
FT   DISULFID    724    738       BY SIMILARITY.
FT   DISULFID    740    749       BY SIMILARITY.
FT   DISULFID    756    767       BY SIMILARITY.
FT   DISULFID    761    776       BY SIMILARITY.
FT   DISULFID    778    787       BY SIMILARITY.
FT   DISULFID    794    805       BY SIMILARITY.
FT   DISULFID    799    814       BY SIMILARITY.
FT   DISULFID    816    825       BY SIMILARITY.
FT   DISULFID    832    843       BY SIMILARITY.
FT   DISULFID    837    854       BY SIMILARITY.
FT   DISULFID    856    865       BY SIMILARITY.
FT   DISULFID    872    883       BY SIMILARITY.
FT   DISULFID    877    892       BY SIMILARITY.
FT   DISULFID    894    903       BY SIMILARITY.
FT   DISULFID    910    921       BY SIMILARITY.
FT   DISULFID    915    930       BY SIMILARITY.
FT   DISULFID    932    941       BY SIMILARITY.
FT   DISULFID    986    997       BY SIMILARITY.
FT   DISULFID    991   1006       BY SIMILARITY.
FT   DISULFID   1008   1017       BY SIMILARITY.
FT   DISULFID   1024   1035       BY SIMILARITY.
FT   DISULFID   1029   1044       BY SIMILARITY.
FT   DISULFID   1046   1055       BY SIMILARITY.
FT   DISULFID   1062   1073       BY SIMILARITY.
FT   DISULFID   1067   1082       BY SIMILARITY.
FT   DISULFID   1084   1093       BY SIMILARITY.
FT   DISULFID   1100   1121       BY SIMILARITY.
FT   DISULFID   1115   1130       BY SIMILARITY.
FT   DISULFID   1132   1141       BY SIMILARITY.
FT   DISULFID   1148   1159       BY SIMILARITY.
FT   DISULFID   1153   1168       BY SIMILARITY.
FT   DISULFID   1170   1179       BY SIMILARITY.
FT   DISULFID   1186   1197       BY SIMILARITY.
FT   DISULFID   1191   1206       BY SIMILARITY.
FT   DISULFID   1208   1217       BY SIMILARITY.
FT   DISULFID   1224   1243       BY SIMILARITY.
FT   DISULFID   1237   1252       BY SIMILARITY.
FT   DISULFID   1254   1263       BY SIMILARITY.
FT   DISULFID   1270   1283       BY SIMILARITY.
FT   DISULFID   1275   1292       BY SIMILARITY.
FT   DISULFID   1294   1303       BY SIMILARITY.
FT   DISULFID   1310   1321       BY SIMILARITY.
FT   DISULFID   1315   1333       BY SIMILARITY.
FT   DISULFID   1335   1344       BY SIMILARITY.
FT   DISULFID   1351   1362       BY SIMILARITY.
FT   DISULFID   1356   1371       BY SIMILARITY.
FT   DISULFID   1373   1382       BY SIMILARITY.
FT   DISULFID   1390   1401       BY SIMILARITY.
FT   DISULFID   1395   1412       BY SIMILARITY.
FT   DISULFID   1414   1423       BY SIMILARITY.
FT   CARBOHYD    462    462       POTENTIAL.
FT   CARBOHYD    887    887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          25.4%;  Score 205;  DB 1;  Length 2524;
  Best Local Similarity  42.5%;  Pred. No. 2.33e-29;
  Matches   34;  Conservative   16;  Mismatches  23;  Indels   7;  Gaps   4;

Db    946 NECASNPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCT-ESS-C--F---NGGTCI 998
          ::| :: |::|| | | ||:||| |  ||||: ||   |    ::| |  :   ||: ||
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    999 DGINTFTCQCPPGFTGSYCQ 1018
              ||:|||||:|: |:
```

```
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT    7
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643       EXTRACELLULAR.
FT   TRANSMEM  1644   1664       POTENTIAL.
FT   DOMAIN    1665   2318       CYTOPLASMIC.
FT   DOMAIN      39   1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261       PEST.
FT   DOMAIN      39     78       EGF-LIKE 1.
FT   DOMAIN      79    119       EGF-LIKE 2.
FT   DOMAIN     120    157       EGF-LIKE 3.
FT   DOMAIN     159    196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235       EGF-LIKE 5.
FT   DOMAIN     237    273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313       EGF-LIKE 7.
FT   DOMAIN     315    351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390       EGF-LIKE 9.
FT   DOMAIN     392    430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731       EGF-LIKE 18.
FT   DOMAIN     735    771       EGF-LIKE 19.
FT   DOMAIN     772    809       EGF-LIKE 20.
FT   DOMAIN     811    848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961       EGF-LIKE 24.
FT   DOMAIN     963    999       EGF-LIKE 25.
FT   DOMAIN    1001   1035       EGF-LIKE 26.
FT   DOMAIN    1037   1083       EGF-LIKE 27.
FT   DOMAIN    1085   1121       EGF-LIKE 28.
FT   DOMAIN    1123   1159       EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245       EGF-LIKE 31.
FT   DOMAIN    1247   1288       EGF-LIKE 32.
FT   DOMAIN    1290   1326       EGF-LIKE 33.
FT   DOMAIN    1336   1374       EGF-LIKE 34.
FT   REPEAT    1388   1428       LIN/NOTCH 1.
FT   REPEAT    1429   1467       LIN/NOTCH 2.
FT   REPEAT    1468   1503       LIN/NOTCH 3.
FT   REPEAT    1784   1816       CDC10/SWI6 1.
FT   REPEAT    1817   1865       CDC10/SWI6 2.
FT   REPEAT    1866   1898       CDC10/SWI6 3.
FT   REPEAT    1899   1932       CDC10/SWI6 4.
FT   REPEAT    1933   1965       CDC10/SWI6 5.
FT   REPEAT    1966   1998       CDC10/SWI6 6.
FT   DISULFID    43     55       BY SIMILARITY.
FT   DISULFID    49     66       BY SIMILARITY.
FT   DISULFID    68     77       BY SIMILARITY.
FT   DISULFID    83     94       BY SIMILARITY.
FT   DISULFID    88    107       BY SIMILARITY.
FT   DISULFID   109    118       BY SIMILARITY.
FT   DISULFID   124    135       BY SIMILARITY.
FT   DISULFID   129    145       BY SIMILARITY.
FT   DISULFID   147    156       BY SIMILARITY.
FT   DISULFID   163    175       BY SIMILARITY.
FT   DISULFID   169    184       BY SIMILARITY.
FT   DISULFID   186    195       BY SIMILARITY.
FT   DISULFID   202    213       BY SIMILARITY.
FT   DISULFID   207    223       BY SIMILARITY.
FT   DISULFID   225    234       BY SIMILARITY.
FT   DISULFID   241    252       BY SIMILARITY.
FT   DISULFID   246    261       BY SIMILARITY.
FT   DISULFID   263    272       BY SIMILARITY.
FT   DISULFID   279    292       BY SIMILARITY.
FT   DISULFID   286    301       BY SIMILARITY.
FT   DISULFID   303    312       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   324    339       BY SIMILARITY.
FT   DISULFID   341    350       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    378       BY SIMILARITY.
FT   DISULFID   380    389       BY SIMILARITY.
FT   DISULFID   396    409       BY SIMILARITY.
FT   DISULFID   403    418       BY SIMILARITY.
FT   DISULFID   420    429       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   441    456       BY SIMILARITY.
FT   DISULFID   458    467       BY SIMILARITY.
FT   DISULFID   474    485       BY SIMILARITY.
FT   DISULFID   479    494       BY SIMILARITY.
FT   DISULFID   496    505       BY SIMILARITY.
FT   DISULFID   512    523       BY SIMILARITY.
FT   DISULFID   517    532       BY SIMILARITY.
FT   DISULFID   534    543       BY SIMILARITY.
FT   DISULFID   550    560       BY SIMILARITY.
FT   DISULFID   555    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   587    598       BY SIMILARITY.
FT   DISULFID   592    607       BY SIMILARITY.
```

```
FT   DISULFID    609    618       BY SIMILARITY.
FT   DISULFID    625    635       BY SIMILARITY.
FT   DISULFID    630    644       BY SIMILARITY.
FT   DISULFID    646    655       BY SIMILARITY.
FT   DISULFID    662    673       BY SIMILARITY.
FT   DISULFID    667    682       BY SIMILARITY.
FT   DISULFID    684    693       BY SIMILARITY.
FT   DISULFID    700    710       BY SIMILARITY.
FT   DISULFID    705    719       BY SIMILARITY.
FT   DISULFID    721    730       BY SIMILARITY.
FT   DISULFID    739    750       BY SIMILARITY.
FT   DISULFID    744    759       BY SIMILARITY.
FT   DISULFID    761    770       BY SIMILARITY.
FT   DISULFID    776    787       BY SIMILARITY.
FT   DISULFID    781    797       BY SIMILARITY.
FT   DISULFID    799    808       BY SIMILARITY.
FT   DISULFID    815    827       BY SIMILARITY.
FT   DISULFID    821    836       BY SIMILARITY.
FT   DISULFID    838    847       BY SIMILARITY.
FT   DISULFID    854    865       BY SIMILARITY.
FT   DISULFID    859    874       BY SIMILARITY.
FT   DISULFID    876    885       BY SIMILARITY.
FT   DISULFID    892    902       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    913    922       BY SIMILARITY.
FT   DISULFID    929    940       BY SIMILARITY.
FT   DISULFID    934    949       BY SIMILARITY.
FT   DISULFID    951    960       BY SIMILARITY.
FT   DISULFID    967    978       BY SIMILARITY.
FT   DISULFID    972    987       BY SIMILARITY.
FT   DISULFID    989    998       BY SIMILARITY.
FT   DISULFID   1005   1016       BY SIMILARITY.
FT   DISULFID   1010   1023       BY SIMILARITY.
FT   DISULFID   1025   1034       BY SIMILARITY.
FT   DISULFID   1041   1062       BY SIMILARITY.
FT   DISULFID   1056   1071       BY SIMILARITY.
FT   DISULFID   1073   1082       BY SIMILARITY.
FT   DISULFID   1089   1100       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          25.0%;  Score 202;  DB 1;  Length 2318;
  Best Local Similarity  39.8%;  Pred. No. 1.32e-28;
  Matches   33;  Conservative   19;  Mismatches  23;  Indels   8;  Gaps   5;

Db    510 DECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCERN---VD-DCSP-DP--CHHG-RCV 561
          |:| :  ||:|| |||: :|| | |::|| | :||:    |  : || |   |::| :|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    562 DGIASFSCACAPGYTGIRCESQV 584
                :| |:||::| |||  :
Qy     63 VVQQEPTCRCPPGFAGPRCEKLI 85
```

```
RESULT    8
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
```

```
FT   REPEAT     1978   2009       ANK MOTIF 4.
FT   REPEAT     2011   2042       ANK MOTIF 5.
FT   REPEAT     2044   2075       ANK MOTIF 6.
FT   CARBOHYD    888    888       POTENTIAL.
FT   CARBOHYD    959    959       POTENTIAL.
FT   CARBOHYD   1179   1179       POTENTIAL.
FT   CARBOHYD   1241   1241       POTENTIAL.
FT   CARBOHYD   1489   1489       POTENTIAL.
FT   CARBOHYD   1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match           24.5%;  Score 198;  DB 1;  Length 2531;
  Best Local Similarity  40.0%;  Pred. No. 1.33e-27;
  Matches   32;  Conservative   19;  Mismatches  22;  Indels   7;  Gaps   4;

Db    947 NECASNPCQNGANCTDCVDSYTCTCPVGFNGIHCENNTPDCT-ESS-C--F---NGGTCV 999
          ::| :: |::|| | | |::||| || ||:|: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINSFTCLCPPGFTGSYCQ 1019
               || |||||:|: |:
     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT    9
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
FT   DOMAIN     869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     907    943       EGF-LIKE 24.
FT   DOMAIN     945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     983   1019       EGF-LIKE 26.
FT   DOMAIN    1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1059   1095       EGF-LIKE 28.
FT   DOMAIN    1097   1143       EGF-LIKE 29.
FT   DOMAIN    1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1267   1305       EGF-LIKE 33.
FT   DOMAIN    1307   1346       EGF-LIKE 34.
FT   DOMAIN    1348   1384       EGF-LIKE 35.
FT   DOMAIN    1387   1426       EGF-LIKE 36.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1865   2076       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2076       ANK MOTIF 6.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
FT   DISULFID   245    254       BY SIMILARITY.
FT   DISULFID   261    272       BY SIMILARITY.
FT   DISULFID   266    281       BY SIMILARITY.
FT   DISULFID   283    292       BY SIMILARITY.
FT   DISULFID   299    312       BY SIMILARITY.
FT   DISULFID   306    321       BY SIMILARITY.
FT   DISULFID   323    332       BY SIMILARITY.
```

```
FT   DISULFID   339   350       BY SIMILARITY.
FT   DISULFID   344   359       BY SIMILARITY.
FT   DISULFID   361   370       BY SIMILARITY.
FT   DISULFID   376   387       BY SIMILARITY.
FT   DISULFID   381   398       BY SIMILARITY.
FT   DISULFID   400   409       BY SIMILARITY.
FT   DISULFID   416   429       BY SIMILARITY.
FT   DISULFID   423   438       BY SIMILARITY.
FT   DISULFID   440   449       BY SIMILARITY.
FT   DISULFID   456   467       BY SIMILARITY.
FT   DISULFID   461   476       BY SIMILARITY.
FT   DISULFID   478   487       BY SIMILARITY.
FT   DISULFID   494   505       BY SIMILARITY.
FT   DISULFID   499   514       BY SIMILARITY.
FT   DISULFID   516   525       BY SIMILARITY.
FT   DISULFID   532   543       BY SIMILARITY.
FT   DISULFID   537   552       BY SIMILARITY.
FT   DISULFID   554   563       BY SIMILARITY.
FT   DISULFID   570   580       BY SIMILARITY.
FT   DISULFID   575   589       BY SIMILARITY.
FT   DISULFID   591   600       BY SIMILARITY.
FT   DISULFID   607   618       BY SIMILARITY.
FT   DISULFID   612   627       BY SIMILARITY.
FT   DISULFID   629   638       BY SIMILARITY.
FT   DISULFID   645   655       BY SIMILARITY.
FT   DISULFID   650   664       BY SIMILARITY.
FT   DISULFID   666   675       BY SIMILARITY.
FT   DISULFID   682   693       BY SIMILARITY.
FT   DISULFID   687   702       BY SIMILARITY.
FT   DISULFID   704   713       BY SIMILARITY.
FT   DISULFID   720   730       BY SIMILARITY.
FT   DISULFID   725   739       BY SIMILARITY.
FT   DISULFID   741   750       BY SIMILARITY.
FT   DISULFID   757   768       BY SIMILARITY.
FT   DISULFID   762   777       BY SIMILARITY.
FT   DISULFID   779   788       BY SIMILARITY.
FT   DISULFID   795   806       BY SIMILARITY.
FT   DISULFID   800   815       BY SIMILARITY.
FT   DISULFID   817   826       BY SIMILARITY.
FT   DISULFID   833   844       BY SIMILARITY.
FT   DISULFID   838   855       BY SIMILARITY.
FT   DISULFID   857   866       BY SIMILARITY.
FT   DISULFID   873   884       BY SIMILARITY.
FT   DISULFID   878   893       BY SIMILARITY.
FT   DISULFID   895   904       BY SIMILARITY.
FT   DISULFID   911   922       BY SIMILARITY.
FT   DISULFID   916   931       BY SIMILARITY.
FT   DISULFID   933   942       BY SIMILARITY.
FT   DISULFID   987   998       BY SIMILARITY.
FT   DISULFID   992  1007       BY SIMILARITY.
FT   DISULFID  1009  1018       BY SIMILARITY.
FT   DISULFID  1025  1036       BY SIMILARITY.
FT   DISULFID  1030  1045       BY SIMILARITY.
FT   DISULFID  1047  1056       BY SIMILARITY.
FT   DISULFID  1063  1074       BY SIMILARITY.
FT   DISULFID  1068  1083       BY SIMILARITY.
FT   DISULFID  1085  1094       BY SIMILARITY.
FT   DISULFID  1101  1122       BY SIMILARITY.
FT   DISULFID  1116  1131       BY SIMILARITY.
FT   DISULFID  1133  1142       BY SIMILARITY.
FT   DISULFID  1149  1160       BY SIMILARITY.
FT   DISULFID  1154  1169       BY SIMILARITY.
FT   DISULFID  1171  1180       BY SIMILARITY.
FT   DISULFID  1187  1198       BY SIMILARITY.
FT   DISULFID  1192  1207       BY SIMILARITY.
FT   DISULFID  1209  1218       BY SIMILARITY.
FT   DISULFID  1225  1244       BY SIMILARITY.
FT   DISULFID  1238  1253       BY SIMILARITY.
FT   DISULFID  1255  1264       BY SIMILARITY.
FT   DISULFID  1271  1284       BY SIMILARITY.
FT   DISULFID  1276  1293       BY SIMILARITY.
FT   DISULFID  1295  1304       BY SIMILARITY.
FT   DISULFID  1311  1322       BY SIMILARITY.
FT   DISULFID  1316  1334       BY SIMILARITY.
FT   DISULFID  1336  1345       BY SIMILARITY.
FT   DISULFID  1352  1363       BY SIMILARITY.
FT   DISULFID  1357  1372       BY SIMILARITY.
FT   DISULFID  1374  1383       BY SIMILARITY.
FT   DISULFID  1391  1403       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          24.5%;  Score 198;  DB 1;  Length 2531;
  Best Local Similarity  40.0%;  Pred. No. 1.33e-27;
  Matches   32;  Conservative   18;  Mismatches  23;  Indels   7;  Gaps   4;

Db    947 NECATNPCQNGANCTDCVDSYTCTCPTGFNGIHCENNTPDCT-ESS-C--F---NGGTCV 999
          ::|  : |::|| | | |::||| || ||:|: ||:  |    ::| |  :   ||: |:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db   1000 DGINSFTCLCPPGFTGSYCQ 1019
               || |||||:|: |:
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT  10
ID   LI12_CAEEL     STANDARD;      PRT;  1429 AA.
AC   P14585;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   LIN-12 PROTEIN PRECURSOR.
GN   LIN-12 OR R107.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 88334747.
RA   YOCHEM J., WESTON K., GREENWALD I.;
RT   "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
RT   protein with overall similarity to Drosophila Notch.";
RL   NATURE 335:547-550(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
CC   -!- FUNCTION: LIN-12 IS IS INVOLVED IN SEVERAL CELL FATES DECISIONS
CC       THAT REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
CC       ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
CC       EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
CC   -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12069; G156358; -.
DR   EMBL; Z14092; E1348691; -.
DR   PIR; S06434; S06434.
DR   WORMPEP; R107.8; CE00274.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 12.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 13.
DR   PFAM; PF00023; ank; 4.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1     15       POTENTIAL.
FT   CHAIN       16   1429       LIN-12 PROTEIN.
FT   DOMAIN      16    908       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   909    931       POTENTIAL.
FT   DOMAIN     932   1429       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24    618       13 X EGF-TYPE REPEATS.
FT   DOMAIN     631    750       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1046   1266       6 X ANK MOTIF REPEATS.
FT   DOMAIN      20     61       EGF-LIKE 1.
FT   DOMAIN     114    150       EGF-LIKE 2.
FT   DOMAIN     152    190       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     201    246       EGF-LIKE 4.
FT   DOMAIN     250    285       EGF-LIKE 5.
FT   DOMAIN     287    323       EGF-LIKE 6.
FT   DOMAIN     323    363       EGF-LIKE 7.
FT   DOMAIN     365    402       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    441       EGF-LIKE 9.
FT   DOMAIN     449    492       EGF-LIKE 10.
FT   DOMAIN     503    541       EGF-LIKE 11.
FT   DOMAIN     543    579       EGF-LIKE 12.
FT   DOMAIN     582    619       EGF-LIKE 13.
FT   REPEAT     635    669       LIN/NOTCH 1.
FT   REPEAT     670    710       LIN/NOTCH 2.
FT   REPEAT     711    750       LIN/NOTCH 3.
FT   REPEAT    1046   1078       ANK MOTIF 1.
FT   REPEAT    1079   1119       ANK MOTIF 2.
FT   REPEAT    1120   1152       ANK MOTIF 3.
FT   REPEAT    1153   1188       ANK MOTIF 4.
FT   REPEAT    1189   1232       ANK MOTIF 5.
FT   REPEAT    1233   1266       ANK MOTIF 6.
FT   DISULFID    24     35       BY SIMILARITY.
FT   DISULFID    29     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID   118    129       BY SIMILARITY.
FT   DISULFID   123    138       BY SIMILARITY.
FT   DISULFID   140    149       BY SIMILARITY.
FT   DISULFID   156    169       BY SIMILARITY.
FT   DISULFID   163    178       BY SIMILARITY.
FT   DISULFID   180    189       BY SIMILARITY.
FT   DISULFID   205    227       BY SIMILARITY.
FT   DISULFID   221    234       BY SIMILARITY.
FT   DISULFID   236    245       BY SIMILARITY.
FT   DISULFID   254    264       BY SIMILARITY.
FT   DISULFID   259    273       BY SIMILARITY.
FT   DISULFID   275    284       BY SIMILARITY.
FT   DISULFID   291    302       BY SIMILARITY.
FT   DISULFID   296    311       BY SIMILARITY.
FT   DISULFID   313    322       BY SIMILARITY.
FT   DISULFID   327    339       BY SIMILARITY.
FT   DISULFID   334    351       BY SIMILARITY.
FT   DISULFID   353    362       BY SIMILARITY.
FT   DISULFID   369    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    429       BY SIMILARITY.
FT   DISULFID   431    440       BY SIMILARITY.
FT   DISULFID   507    518       BY SIMILARITY.
FT   DISULFID   512    529       BY SIMILARITY.
FT   DISULFID   531    540       BY SIMILARITY.
FT   DISULFID   547    558       BY SIMILARITY.
FT   DISULFID   552    567       BY SIMILARITY.
FT   DISULFID   569    578       BY SIMILARITY.
FT   DISULFID   586    597       BY SIMILARITY.
FT   DISULFID   591    607       BY SIMILARITY.
FT   DISULFID   609    618       BY SIMILARITY.
FT   CARBOHYD    41     41       POTENTIAL.
FT   CARBOHYD   165    165       POTENTIAL.
FT   CARBOHYD   194    194       POTENTIAL.
FT   CARBOHYD   378    378       POTENTIAL.
FT   CARBOHYD   515    515       POTENTIAL.
FT   CARBOHYD   623    623       POTENTIAL.
FT   CARBOHYD   751    751       POTENTIAL.
FT   CARBOHYD   754    754       POTENTIAL.
FT   CARBOHYD   900    900       POTENTIAL.
SQ   SEQUENCE   1429 AA;  157115 MW;  CFD2CCA4 CRC32;

  Query Match          24.2%;  Score 195;  DB 1;  Length 1429;
  Best Local Similarity  30.1%;  Pred. No. 7.46e-27;
  Matches   28;  Conservative   27;  Mismatches  29;  Indels   9;  Gaps   4;

Db    365 DKNECLSENMCLNNGTCVNLPGSFRCDCARGFGGKWCDEP-----LNM--CQDFHCENDG 417
          ::::|:: ::| : : ||:   :: | |::||:|  |: |     |   |::: |:| :
Qy      1 NNDDCVG-HKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGA 59

Db    418 TCMHTSDHSPVCQCKNGFIGKRCEKECPIGFGG 450
            |:   :: | |:|  || | ||||   : | |
Qy     60 QCI-VVQQEPTCRCPPGFAGPRCEKLITVNFVG 91


RESULT  11
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   570       FIBROPELLIN C.
FT   DOMAIN      18    55       EGF-LIKE 1.
FT   DOMAIN      62   175       CUB.
FT   DOMAIN     176   212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214   250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252   288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290   326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328   364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366   402       EGF-LIKE 7.
FT   DOMAIN     404   440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442   570       AVIDIN-LIKE.
FT   DISULFID    23    34       BY SIMILARITY.
FT   DISULFID    28    43       BY SIMILARITY.
FT   DISULFID    45    54       BY SIMILARITY.
FT   DISULFID   180   191       BY SIMILARITY.
FT   DISULFID   185   200       BY SIMILARITY.
FT   DISULFID   202   211       BY SIMILARITY.
FT   DISULFID   218   229       BY SIMILARITY.
FT   DISULFID   223   238       BY SIMILARITY.
FT   DISULFID   240   249       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   261   276       BY SIMILARITY.
FT   DISULFID   278   287       BY SIMILARITY.
FT   DISULFID   294   305       BY SIMILARITY.
FT   DISULFID   299   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   CARBOHYD    30    30       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   357   357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match          23.9%;  Score 193;  DB 1;  Length 570;
  Best Local Similarity  41.2%;  Pred. No. 2.35e-26;
  Matches   35;  Conservative   12;  Mismatches  31;  Indels   7;  Gaps   5;

Db    176 DGDDCTPNPCLNGATCVDQVNDYQCICAPGFTGDNCE--TD-I--D-E-CASAPCRNGGA 228
          : |||  : | :|| |||:|| | |||: ||:|  ||     :  :   |    |:||:
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    229 CVDQVNGYTCNCIPGFNGVNCENNI 253
          |:      || | ||| |  ||: |
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLI 85


RESULT   12
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   MECH. DEV. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69088; G433867; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1    20       POTENTIAL.
FT   CHAIN       21  2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21  1724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725  1747       POTENTIAL.
FT   DOMAIN    1748  2437       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21    57       EGF-LIKE 1.
FT   DOMAIN      58    98       EGF-LIKE 2.
FT   DOMAIN     101   138       EGF-LIKE 3.
FT   DOMAIN     139   175       EGF-LIKE 4.
FT   DOMAIN     177   215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217   254       EGF-LIKE 6.
FT   DOMAIN     256   292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294   332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334   370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371   409       EGF-LIKE 10.
FT   DOMAIN     411   449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451   487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489   524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526   562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564   599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     601   637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      639    674      EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      676    712      EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      714    749      EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      751    787      EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      789    825      EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      827    865      EGF-LIKE 22.
FT   DOMAIN      867    903      EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      905    941      EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      943    979      EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      981   1017      EGF-LIKE 26.
FT   DOMAIN     1019   1055      EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1057   1093      EGF-LIKE 28.
FT   DOMAIN     1095   1141      EGF-LIKE 29.
FT   DOMAIN     1143   1179      EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1181   1217      EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1219   1263      EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1265   1303      EGF-LIKE 33.
FT   DOMAIN     1305   1344      EGF-LIKE 34.
FT   DOMAIN     1346   1382      EGF-LIKE 35.
FT   DOMAIN     1385   1423      EGF-LIKE 36.
FT   DOMAIN     1446   1561      3 X LIN/NOTCH REPEATS.
FT   REPEAT     1446   1486      LIN/NOTCH 1.
FT   REPEAT     1487   1520      LIN/NOTCH 2.
FT   REPEAT     1521   1561      LIN/NOTCH 3.
FT   DOMAIN     1861   2074      6 X ANK MOTIF REPEATS.
FT   REPEAT     1861   1891      ANK MOTIF 1.
FT   REPEAT     1892   1940      ANK MOTIF 1.
FT   REPEAT     1941   1974      ANK MOTIF 1.
FT   REPEAT     1975   2007      ANK MOTIF 1.
FT   REPEAT     2008   2040      ANK MOTIF 1.
FT   REPEAT     2041   2074      ANK MOTIF 1.
FT   DOMAIN     2265   2276      POLY-GLN (OPA-REPEAT).
FT   DISULFID     25     35      BY SIMILARITY.
FT   DISULFID     29     45      BY SIMILARITY.
FT   DISULFID     47     56      BY SIMILARITY.
FT   DISULFID     62     73      BY SIMILARITY.
FT   DISULFID     67     86      BY SIMILARITY.
FT   DISULFID     88     97      BY SIMILARITY.
FT   DISULFID    105    116      BY SIMILARITY.
FT   DISULFID    110    126      BY SIMILARITY.
FT   DISULFID    128    137      BY SIMILARITY.
FT   DISULFID    143    154      BY SIMILARITY.
FT   DISULFID    148    163      BY SIMILARITY.
FT   DISULFID    165    174      BY SIMILARITY.
FT   DISULFID    181    194      BY SIMILARITY.
FT   DISULFID    188    203      BY SIMILARITY.
FT   DISULFID    205    214      BY SIMILARITY.
FT   DISULFID    221    232      BY SIMILARITY.
FT   DISULFID    226    242      BY SIMILARITY.
FT   DISULFID    244    253      BY SIMILARITY.
FT   DISULFID    260    271      BY SIMILARITY.
FT   DISULFID    265    280      BY SIMILARITY.
FT   DISULFID    282    291      BY SIMILARITY.
FT   DISULFID    298    311      BY SIMILARITY.
FT   DISULFID    305    320      BY SIMILARITY.
FT   DISULFID    322    331      BY SIMILARITY.
FT   DISULFID    338    349      BY SIMILARITY.
FT   DISULFID    343    358      BY SIMILARITY.
FT   DISULFID    360    369      BY SIMILARITY.
FT   DISULFID    375    386      BY SIMILARITY.
FT   DISULFID    380    397      BY SIMILARITY.
FT   DISULFID    399    408      BY SIMILARITY.
FT   DISULFID    415    428      BY SIMILARITY.
FT   DISULFID    422    437      BY SIMILARITY.
FT   DISULFID    439    448      BY SIMILARITY.
FT   DISULFID    455    466      BY SIMILARITY.
FT   DISULFID    460    475      BY SIMILARITY.
FT   DISULFID    477    486      BY SIMILARITY.
FT   DISULFID    493    503      BY SIMILARITY.
FT   DISULFID    498    512      BY SIMILARITY.
FT   DISULFID    514    523      BY SIMILARITY.
FT   DISULFID    530    541      BY SIMILARITY.
FT   DISULFID    535    550      BY SIMILARITY.
FT   DISULFID    552    561      BY SIMILARITY.
FT   DISULFID    568    578      BY SIMILARITY.
FT   DISULFID    573    587      BY SIMILARITY.
FT   DISULFID    589    598      BY SIMILARITY.
FT   DISULFID    605    616      BY SIMILARITY.
FT   DISULFID    610    625      BY SIMILARITY.
FT   DISULFID    627    636      BY SIMILARITY.
FT   DISULFID    643    653      BY SIMILARITY.
FT   DISULFID    648    662      BY SIMILARITY.
FT   DISULFID    664    673      BY SIMILARITY.
FT   DISULFID    680    691      BY SIMILARITY.
FT   DISULFID    685    700      BY SIMILARITY.
FT   DISULFID    702    711      BY SIMILARITY.
FT   DISULFID    718    728      BY SIMILARITY.
FT   DISULFID    723    737      BY SIMILARITY.
FT   DISULFID    739    748      BY SIMILARITY.
FT   DISULFID    755    766      BY SIMILARITY.
FT   DISULFID    760    775      BY SIMILARITY.
FT   DISULFID    777    786      BY SIMILARITY.
FT   DISULFID    793    804      BY SIMILARITY.
FT   DISULFID    798    813      BY SIMILARITY.
FT   DISULFID    815    824      BY SIMILARITY.
FT   DISULFID    831    842      BY SIMILARITY.
FT   DISULFID    836    853      BY SIMILARITY.
FT   DISULFID    855    864      BY SIMILARITY.
FT   DISULFID    871    882      BY SIMILARITY.
FT   DISULFID    876    891      BY SIMILARITY.
FT   DISULFID    893    902      BY SIMILARITY.
FT   DISULFID    909    920      BY SIMILARITY.
FT   DISULFID    914    929      BY SIMILARITY.
FT   DISULFID    931    940      BY SIMILARITY.
FT   DISULFID    947    958      BY SIMILARITY.
FT   DISULFID    952    967      BY SIMILARITY.
FT   DISULFID    969    978      BY SIMILARITY.
FT   DISULFID   1023   1034      BY SIMILARITY.
FT   DISULFID   1028   1043      BY SIMILARITY.
FT   DISULFID   1045   1054      BY SIMILARITY.
FT   DISULFID   1061   1072      BY SIMILARITY.
FT   DISULFID   1066   1081      BY SIMILARITY.
FT   DISULFID   1083   1092      BY SIMILARITY.
FT   DISULFID   1099   1120      BY SIMILARITY.
FT   DISULFID   1114   1129      BY SIMILARITY.
FT   DISULFID   1131   1140      BY SIMILARITY.
FT   DISULFID   1147   1158      BY SIMILARITY.
FT   DISULFID   1152   1167      BY SIMILARITY.
FT   DISULFID   1169   1178      BY SIMILARITY.
FT   DISULFID   1185   1196      BY SIMILARITY.
FT   DISULFID   1190   1205      BY SIMILARITY.
FT   DISULFID   1207   1216      BY SIMILARITY.
FT   DISULFID   1223   1242      BY SIMILARITY.
FT   DISULFID   1236   1251      BY SIMILARITY.
FT   DISULFID   1253   1262      BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          23.8%;  Score 192;  DB 1;  Length 2437;
  Best Local Similarity  35.8%;  Pred. No. 4.16e-26;
  Matches   29;  Conservative   16;  Mismatches  35;  Indels   1;  Gaps   1;

Db   1183 NECLSQPCQNGGTCIDLVNTYKCSCPRGTQGVHCEIDIDDCSPSVDPLTGEPRCFNGGRC 1242
            ::|::: |::|: |:| || | | ||:|  |: ||          |   :  | ||::|
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSP-CDQYECQNGAQC 61

Db   1243 VDRVGGYGCVCPAGFVGERCE 1263
            :       | ||:|| | |||
Qy     62 IVVQQEPTCRCPPGFAGPRCE 82


RESULT   13
ID   SERR_DROME     STANDARD;      PRT;  1408 AA.
AC   P18168;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
```

```
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SERRATE PROTEIN PRECURSOR (BEADED PROTEIN).
GN   SER OR BD.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 91347903.
RA   THOMAS U., SPEICHER S.A., KNUST E.;
RT   "The Drosophila gene Serrate encodes an EGF-like transmembrane
RT   protein with a complex expression pattern in embryos and wing
RT   discs.";
RL   DEVELOPMENT 111:749-761(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099666.
RA   FLEMING R.J., SCOTTGALE T.N., DIEDERICH R.J., ARTAVANIS-TSAKONAS S.;
RT   "The gene Serrate encodes a putative EGF-like transmembrane protein
RT   essential for proper ectodermal development in Drosophila
RT   melanogaster.";
RL   GENES DEV. 4:2188-2201(1990).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER ECTODERMAL DEVELOPMENT. SERRATE
CC       MAY REPRESENT AN ELEMENT IN A NETWORK OF INTERACTING MOLECULES
CC       OPERATING AT THE CELL SURFACE DURING THE DIFFERENTIATION OF
CC       CERTAIN TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: APPEARS TO BE RESTRICTED EXCLUSIVELY TO
CC       CELLS OF ECTODERMAL ORIGIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN LEVEL, IT IS
CC       CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY COMPETE
CC       FOR BINDING WITH THE NOTCH PROTEIN.
CC   -!- SIMILARITY: CONTAINS 14 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO THE DROSOPHILA NEUROGENIC LOCUS DELTA PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56811; G8564; -.
DR   EMBL; M35759; G158606; -.
DR   PIR; A36666; A36666.
DR   PIR; S16878; S16878.
DR   FLYBASE; FBgn0004197; Ser.
DR   PROSITE; PS00010; ASX_HYDROXYL; 7.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 11.
DR   HSSP; P00743; 1WHE.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1     83       POTENTIAL.
FT   CHAIN       84   1408       SERRATE PROTEIN.
FT   DOMAIN      84   1223       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1224   1249       POTENTIAL.
FT   DOMAIN    1250   1408       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     284    317       EGF-LIKE 1.
FT   DOMAIN     315    349       EGF-LIKE 2.
FT   DOMAIN     351    389       EGF-LIKE 3.
FT   DOMAIN     391    489       EGF-LIKE 4.
FT   DOMAIN     407    476       SER-RICH (INSERT).
FT   DOMAIN     491    527       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     529    609       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     611    646       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     648    684       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     686    721       EGF-LIKE 9.
FT   DOMAIN     723    797       EGF-LIKE 10.
FT   DOMAIN     737    769       THR-RICH (INSERT).
FT   DOMAIN     799    835       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     837    877       EGF-LIKE 12.
FT   DOMAIN     879    915       EGF-LIKE 13.
FT   DOMAIN     917    953       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID   288    299       BY SIMILARITY.
FT   DISULFID   292    305       BY SIMILARITY.
FT   DISULFID   307    316       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   325    337       BY SIMILARITY.
FT   DISULFID   339    348       BY SIMILARITY.
FT   DISULFID   355    367       BY SIMILARITY.
FT   DISULFID   361    377       BY SIMILARITY.
FT   DISULFID   379    388       BY SIMILARITY.
FT   DISULFID   395    406       BY SIMILARITY.
FT   DISULFID   400    477       BY SIMILARITY.
FT   DISULFID   479    488       BY SIMILARITY.
FT   DISULFID   495    506       BY SIMILARITY.
FT   DISULFID   500    515       BY SIMILARITY.
FT   DISULFID   517    526       BY SIMILARITY.
FT   DISULFID   533    588       BY SIMILARITY.
FT   DISULFID   582    597       BY SIMILARITY.
FT   DISULFID   599    608       BY SIMILARITY.
FT   DISULFID   615    625       BY SIMILARITY.
FT   DISULFID   619    634       BY SIMILARITY.
FT   DISULFID   636    645       BY SIMILARITY.
FT   DISULFID   652    663       BY SIMILARITY.
FT   DISULFID   657    672       BY SIMILARITY.
FT   DISULFID   674    683       BY SIMILARITY.
FT   DISULFID   690    700       BY SIMILARITY.
FT   DISULFID   695    709       BY SIMILARITY.
FT   DISULFID   711    720       BY SIMILARITY.
FT   DISULFID   803    814       BY SIMILARITY.
FT   DISULFID   808    823       BY SIMILARITY.
FT   DISULFID   825    834       BY SIMILARITY.
FT   DISULFID   841    852       BY SIMILARITY.
FT   DISULFID   846    865       BY SIMILARITY.
FT   DISULFID   867    876       BY SIMILARITY.
FT   DISULFID   883    894       BY SIMILARITY.
FT   DISULFID   888    903       BY SIMILARITY.
FT   DISULFID   905    914       BY SIMILARITY.
FT   DISULFID   921    932       BY SIMILARITY.
FT   DISULFID   926    941       BY SIMILARITY.
FT   DISULFID   943    952       BY SIMILARITY.
FT   CARBOHYD   152    152       POTENTIAL.
FT   CARBOHYD   196    196       POTENTIAL.
FT   CARBOHYD   247    247       POTENTIAL.
FT   CARBOHYD   331    331       POTENTIAL.
FT   CARBOHYD   412    412       POTENTIAL.
FT   CARBOHYD   452    452       POTENTIAL.
FT   CARBOHYD   558    558       POTENTIAL.
FT   CARBOHYD   739    739       POTENTIAL.
FT   CARBOHYD   965    965       POTENTIAL.
FT   CARBOHYD   977    977       POTENTIAL.
FT   CARBOHYD  1004   1004       POTENTIAL.
FT   CARBOHYD  1030   1030       POTENTIAL.
FT   CARBOHYD  1150   1150       POTENTIAL.
FT   CONFLICT    14     17       MISSING (IN REF. 2).
FT   CONFLICT    27     27       P -> A (IN REF. 2).
FT   CONFLICT  1352   1352       T -> S (IN REF. 2).
SQ   SEQUENCE  1408 AA;  150660 MW;  A494A358 CRC32;

  Query Match          23.5%;  Score 190;  DB 1;  Length 1408;
  Best Local Similarity  41.3%;  Pred. No. 1.30e-25;
  Matches   33;  Conservative   15;  Mismatches   24;  Indels   8;  Gaps   6;

Db    613 DDCVGQ-CRNGATCIDLVNDYRCACASGFTGRDCE--TD-I--D-E-CATSPCRNGGECV 664
          |||||: ||:|| |:| || | | |: ||:|  ||     :  :   |    |:||::|:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62
```

```
Db    665 DMVGKFNCICPLGYSGSLCE 684
          :     | || |::|: ||
Qy     63 VVQQEPTCRCPPGFAGPRCE 82


RESULT   14
ID   DLL1_MOUSE     STANDARD;      PRT;   722 AA.
AC   Q61483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C X C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95401858.
RA   BETTENHAUSEN B., DE ANGELIS M.H., SIMON D., GUENET J.-L., GOSSLER A.;
RT   "Transient and restricted expression during mouse embryogenesis of
RT   Dll1, a murine gene closely related to Drosophila Delta.";
RL   DEVELOPMENT 121:2407-2418(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN THE PARAXIAL
CC       MESODERM AND NERVOUS SYSTEM. EXPRESSED AT HIGH LEVELS IN ADULT
CC       HEART AND AT LOWER LEVELS, IN ADULT LUNG.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED UNTIL DAY 15 IN THE EMBRYO.
CC       EXPRESSION THEN DECREASES AND INCREASES AGAIN IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X80903; G806570; -.
DR   MGD; MGI:104659; DLL1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   722       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   722       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     225   253       EGF-LIKE 1.
FT   DOMAIN     256   284       EGF-LIKE 2.
FT   DOMAIN     291   324       EGF-LIKE 3.
FT   DOMAIN     331   362       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     369   401       EGF-LIKE 5.
FT   DOMAIN     408   439       EGF-LIKE 6.
FT   DOMAIN     446   477       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     484   515       EGF-LIKE 8.
FT   DISULFID   225   236       BY SIMILARITY.
FT   DISULFID   229   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   256   267       BY SIMILARITY.
FT   DISULFID   262   273       BY SIMILARITY.
FT   DISULFID   275   284       BY SIMILARITY.
FT   DISULFID   291   303       BY SIMILARITY.
FT   DISULFID   297   313       BY SIMILARITY.
FT   DISULFID   315   324       BY SIMILARITY.
FT   DISULFID   331   342       BY SIMILARITY.
FT   DISULFID   336   351       BY SIMILARITY.
FT   DISULFID   353   362       BY SIMILARITY.
FT   DISULFID   369   380       BY SIMILARITY.
FT   DISULFID   374   390       BY SIMILARITY.
FT   DISULFID   392   401       BY SIMILARITY.
FT   DISULFID   408   419       BY SIMILARITY.
FT   DISULFID   413   428       BY SIMILARITY.
FT   DISULFID   430   439       BY SIMILARITY.
FT   DISULFID   446   466       BY SIMILARITY.
FT   DISULFID   468   477       BY SIMILARITY.
FT   DISULFID   484   495       BY SIMILARITY.
FT   DISULFID   489   504       BY SIMILARITY.
FT   DISULFID   506   515       BY SIMILARITY.
FT   CARBOHYD   476   476       POTENTIAL.
SQ   SEQUENCE   722 AA;  78448 MW;  5A647702 CRC32;

  Query Match          23.3%;  Score 188;  DB 1;  Length 722;
  Best Local Similarity  38.8%;  Pred. No. 4.08e-25;
  Matches   33;  Conservative   17;  Mismatches  28;  Indels   7;  Gaps   3;

Db    442 NVDDCASSPCANGGTCRDSVNDFSCTCPPGYTGKNC--SAP-V----SRCEHAPCHNGAT 494
          | ||| :  | :|: | | || ::| || |::|  |  ::| |    | |::  |:|||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    495 CHQRGQRYMCECAQGYGGPNCQFLL 519
          |    |   | |: |::|| |: |:
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLI 85


RESULT   15
ID   DLL1_HUMAN     STANDARD;      PRT;   723 AA.
AC   O00548;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DELTA-LIKE PROTEIN 1 PRECURSOR (DELTA1).
GN   DLL1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MANN R.S., GRAY G.E., HENRIQUE D., ISH-HOROWICZ D.,
RA   ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-TO-CELL COMMUNICATION IN
CC       MAMMALIAN EMBRYOS. MAY HAVE A ROLE IN CELLULAR INTERACTIONS
CC       UNDERLYING SOMITOGENESIS AND DEVELOPMENT OF THE NERVOUS SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA DELTA PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF003522; G2197069; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 6.
DR   HSSP; P00740; 1IXA.
KW   SIGNAL; EGF-LIKE DOMAIN; GLYCOPROTEIN; TRANSMEMBRANE.
```

```
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   723       DELTA-LIKE PROTEIN 1.
FT   DOMAIN      18   545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   546   568       POTENTIAL.
FT   DOMAIN     569   723       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     226   254       EGF-LIKE 1.
FT   DOMAIN     257   285       EGF-LIKE 2.
FT   DOMAIN     292   325       EGF-LIKE 3.
FT   DOMAIN     332   363       EGF-LIKE 4, CALCIUM BINDING (POTENTIAL).
FT   DOMAIN     370   402       EGF-LIKE 5.
FT   DOMAIN     409   440       EGF-LIKE 6.
FT   DOMAIN     447   478       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     485   516       EGF-LIKE 8.
FT   DISULFID   226   237       BY SIMILARITY.
FT   DISULFID   230   243       BY SIMILARITY.
FT   DISULFID   245   254       BY SIMILARITY.
FT   DISULFID   257   268       BY SIMILARITY.
FT   DISULFID   263   274       BY SIMILARITY.
FT   DISULFID   276   285       BY SIMILARITY.
FT   DISULFID   292   304       BY SIMILARITY.
FT   DISULFID   298   314       BY SIMILARITY.
FT   DISULFID   316   325       BY SIMILARITY.
FT   DISULFID   332   343       BY SIMILARITY.
FT   DISULFID   337   352       BY SIMILARITY.
FT   DISULFID   354   363       BY SIMILARITY.
FT   DISULFID   370   381       BY SIMILARITY.
FT   DISULFID   375   391       BY SIMILARITY.
FT   DISULFID   393   402       BY SIMILARITY.
FT   DISULFID   409   420       BY SIMILARITY.
FT   DISULFID   414   429       BY SIMILARITY.
FT   DISULFID   431   440       BY SIMILARITY.
FT   DISULFID   447   467       BY SIMILARITY.
FT   DISULFID   469   478       BY SIMILARITY.
FT   DISULFID   485   496       BY SIMILARITY.
FT   DISULFID   490   505       BY SIMILARITY.
FT   DISULFID   507   516       BY SIMILARITY.
FT   CARBOHYD   477   477       POTENTIAL.
SQ   SEQUENCE  723 AA;  77956 MW;  A1D48BDB CRC32;

  Query Match          23.3%;  Score 188;  DB 1;  Length 723;
  Best Local Similarity  37.6%;  Pred. No. 4.08e-25;
  Matches   32;  Conservative   18;  Mismatches  28;  Indels   7;  Gaps   3;

Db    443 NVDDCASSPCANGGTCRDGVNDFSCTCPPGYTGRNC--SAP-V----SRCEHAPCHNGAT 495
          | ||| :  | :|: | | || ::| || |::|  |  ::| |    | |::  |:|||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    496 CHERGHGYVCECARGYGGPNCQFLL 520
          |    :   | |: |::|| |: |:
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLI 85


Search completed: Fri May 28 09:36:23 1999
Job time : 26 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:36:42 1999; MasPar time 9.96 Seconds
569.941 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-13
Description: (1-104) from US09191647.pep
Perfect Score: 807
Sequence: 1 NNDDCVGHKCRHGAQCVDEV.....ITVNFVGKDSYVELASAKVR 104

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 36.193; Variance 55.571; scale 0.651

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	793	98.3	1523	11	O88280	MEGF5.	2.41e-180
2	770	95.4	739	4	O75094	MEGF5 (FRAGMENT).	3.98e-174
3	394	48.8	1531	11	O88279	MEGF4.	2.54e-74
4	343	42.5	530	5	Q24526	SLIT LOCUS ENCODING A	3.00e-61
5	307	38.0	601	5	Q20204	F40E10.4 PROTEIN (FRAG	3.79e-52
6	236	29.2	2531	5	O16004	NOTCH HOMOLOG.	1.26e-34
7	227	28.1	529	5	Q25058	FIBROPELLIN IA (FRAGME	1.85e-32
8	212	26.3	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	6.99e-29
9	209	25.9	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	3.59e-28
10	208	25.8	752	13	O42374	NOTCH RECEPTOR PROTEIN	6.18e-28
11	205	25.4	406	5	Q25059	FIBROPELLIN III (FRAGM	3.15e-27
12	203	25.2	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	9.30e-27
13	199	24.7	717	13	P87357	DELTAD TRANSMEMBRANE P	8.07e-26
14	199	24.7	802	13	O57462	DELTAA.	8.07e-26
15	197	24.4	387	11	Q06007	NOTCH PROTEIN HOMOLOG	2.37e-25
16	197	24.4	2447	13	O13149	NOTCH 2 (FRAGMENT).	2.37e-25
17	196	24.3	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	4.05e-25
18	196	24.3	1574	11	O88281	MEGF6.	4.05e-25
19	196	24.3	2470	11	O35516	CELL SURFACE PROTEIN.	4.05e-25
20	193	23.9	1476	13	Q90285	PUTATIVE EXTRACELLULAR	2.03e-24
21	191	23.7	2352	5	O61240	HRNOTCH PROTEIN.	5.90e-24
22	190	23.5	1372	5	P91526	SIMILARITY TO MULTIPLE	1.01e-23
23	190	23.5	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	1.01e-23
24	188	23.3	955	4	Q99466	NOTCH4 (FRAGMENT).	2.92e-23
25	188	23.3	1218	4	Q15816	TRANSMEMBRANE PROTEIN	2.92e-23
26	188	23.3	1218	4	O14902	TRANSMEMBRANE PROTEIN	2.92e-23
27	188	23.3	1219	11	Q63722	JAGGED PROTEIN.	2.92e-23
28	188	23.3	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	2.92e-23
29	188	23.3	1999	4	Q99940	NOTCH4.	2.92e-23
30	188	23.3	2003	4	O00306	NOTCH4.	2.92e-23
31	186	23.0	1218	4	O15122	JAGGED1.	8.47e-23
32	185	22.9	1964	11	O35442	NOTCH4.	1.44e-22
33	182	22.6	156	5	Q26661	EPIDERMAL GROWTH FACTO	7.05e-22
34	180	22.3	263	4	Q99734	NOTCH2 TRANSMEMBRANE P	2.03e-21
35	178	22.1	434	11	O55139	JAGGED2 PROTEIN (FRAGM	5.81e-21
36	178	22.1	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	5.81e-21
37	178	22.1	615	13	O57409	DELTAB.	5.81e-21
38	178	22.1	728	13	Q90656	TRANSMEMBRANE PROTEIN	5.81e-21
39	178	22.1	762	13	O42373	NOTCH RECEPTOR PROTEIN	5.81e-21
40	178	22.1	1202	11	P97607	JAGGED2 (FRAGMENT).	5.81e-21
41	176	21.8	721	13	Q91902	X-DELTA-1.	1.66e-20
42	175	21.7	153	4	O75095	MEGF6 (FRAGMENT).	2.80e-20
43	174	21.6	2824	5	P90891	F55H12.3 PROTEIN.	4.73e-20
44	172	21.3	259	5	Q93519	F16B12.2 PROTEIN.	1.34e-19
45	171	21.2	832	5	Q99108	NEUROGENIC LOCUS DELTA	2.26e-19

ALIGNMENTS

```
RESULT   1
ID   O88280      PRELIMINARY;      PRT;  1523 AA.
AC   O88280;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5.
GN   MEGF5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011531; D1033424; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1523 AA;  167767 MW;  2BD845D0 CRC32;

  Query Match          98.3%;  Score 793;  DB 11;  Length 1523;
  Best Local Similarity  97.1%;  Pred. No. 2.41e-180;
  Matches  101;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1074 DNDDCVAHKCRHGAQCVDAVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 1133
           :|||||:|||||||||||| |||||||||||||||||||||||||||||||||||||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db   1134 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 1177
           |||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT   2
ID   O75094      PRELIMINARY;      PRT;   739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
```

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;

  Query Match           95.4%;  Score 770;  DB 4;  Length 739;
  Best Local Similarity  94.2%;  Pred. No. 3.98e-174;
  Matches   98;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db    290 DNDDCVAHKCRHGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVLLQTSPCDQYECQNGAQ 349
          :|||||:|||||||||||| :||||| ||||||| |||||||||||||||||||||||||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    350 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 393
          |||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    3
ID   O88279      PRELIMINARY;      PRT;  1531 AA.
AC   O88279;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4.
GN   MEGF4.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011530; D1033423; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1531 AA;  167497 MW;  5C5EBDF4 CRC32;

  Query Match           48.8%;  Score 394;  DB 11;  Length 1531;
  Best Local Similarity  51.5%;  Pred. No. 2.54e-74;
  Matches   51;  Conservative   24;  Mismatches  20;  Indels   4;  Gaps   3;

Db   1083 NQDDCKDHQCQNGAQCVDEINSYACLCAEGYSGQLCEIPPA-P--RNS-CEGTECQNGAN 1138
          | |||  | |::|||||||:|:|:|:|::|:|| :|| ||:    : | |:  ||||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db   1139 CVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFT 1177
          |:     | |:| |||:|| ||||::|||| :|:|:::::
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELA 99


RESULT    4
ID   Q24526      PRELIMINARY;      PRT;  530 AA.
AC   Q24526;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SLIT LOCUS ENCODING A PROTEIN ASSOCIATED WITH NEURAL DEVELOPMENT WITH
DE   52D EGF HOMOLOGOUS DOMAINS (FRAGMENT).
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 89077533.
RA   ROTHBERG J.M., HARTLEY D.A., WALTHER Z., ARTAVANIS-TSAKONAS S.;
RT   "slit: an EGF-homologous locus of D. melanogaster involved in the
RT   development of the embryonic central nervous system.";
RL   CELL 55:1047-1059(1988).
DR   EMBL; M23543; G514357; -.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
KW   NEUROGENESIS; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER     530    530
SQ   SEQUENCE   530 AA;  59457 MW;  10E5764D CRC32;

  Query Match           42.5%;  Score 343;  DB 5;  Length 530;
  Best Local Similarity  42.2%;  Pred. No. 3.00e-61;
  Matches   46;  Conservative   26;  Mismatches  31;  Indels   6;  Gaps   5;

Db    184 NIDDCQNHMCQNGGTCVDGINDYQCRCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHG 243
          | |||  |:|::|: ||| :| | | ||: ::| :|| |   :|:  |||||:  || :|
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCE-HPP-PMVLLQTSPCDQYECQNG 58

Db    244 V-CFQPNAQGSDYLCRCHPGYTGKWCEYLTSISFVHNNSFVELEPLRTR 291
           |:    | :   ||| ||::| :|| | :::|| ::|:||| : : |
Qy     59 AQCIVVQ-QEPT--CRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVR 104


RESULT    5
ID   Q20204      PRELIMINARY;      PRT;  601 AA.
AC   Q20204;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F40E10.4 PROTEIN (FRAGMENT).
GN   F40E10.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMYE R.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z69792; E1346469; -.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   601 AA;  66669 MW;  F8A72773 CRC32;

  Query Match           38.0%;  Score 307;  DB 5;  Length 601;
  Best Local Similarity  37.5%;  Pred. No. 3.79e-52;
  Matches    39;  Conservative   24;  Mismatches  36;  Indels   5;  Gaps   4;

Db    216 NIDDCKNVECQNGGSCVDGILSYDCLCRPGYAGQYCEIPPMMDMEYQKTDACQQSACGQG 275
          | |||    |::|: ||| : :| |:|  |::| :|| || | :  | | :|:|  |  |
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVL-LQ-TSPCDQYECQNG 58

Db    276 -ECVASQNSSDFTCKCHEGFSGPSCDRQMSVGFKNPGAYLALDP 318
           :|:  |  :  ||:|  ||:|| |:: ::| |    :|: | :
Qy     59 AQCIVVQQEP--TCRCPPGFAGPRCEKLITVNFVGKDSYVELAS 100


RESULT    6
ID   O16004      PRELIMINARY;      PRT;  2531 AA.
AC   O16004;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH HOMOLOG.
OS   LYTECHINUS VARIEGATUS (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; TEMNOPLEUROIDA; TOXOPNEUSTIDAE; LYTECHINUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97454256.
RA   SHERWOOD D.R., MCCLAY D.R.;
RT   "Identification and localization of a sea urchin Notch homologue:
RT   insights into vegetal plate regionalization and Notch receptor
RT   regulation.";
RL   DEVELOPMENT 124:3363-3374(1997).
DR   EMBL; AF000634; G2570351; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 21.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2531 AA;  273982 MW;  BB9C6F3D CRC32;

  Query Match           29.2%;  Score 236;  DB 5;  Length 2531;
  Best Local Similarity  44.0%;  Pred. No. 1.26e-34;
  Matches    37;  Conservative   17;  Mismatches  21;  Indels   9;  Gaps   4;

Db    779 NIDDCVDEPCLNGGICIDEVNSFQCVCPQTFVGLLCE-----T--ERSPCEDNQCQNGAT 831
          | ||||   | :|: |:||||:: |:||| | ||:||       : |||:: :|||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    832 CVYSEDYAGYSCRCTSGFQGNFCD 855
          |:  ::    :||| :|| |  |:
Qy     61 CIVVQQ-EP-TCRCPPGFAGPRCE 82


RESULT    7
ID   Q25058      PRELIMINARY;      PRT;  529 AA.
AC   Q25058;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN IA (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33861; G499686; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 10.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   529 AA;  55543 MW;  6385F322 CRC32;

  Query Match           28.1%;  Score 227;  DB 5;  Length 529;
  Best Local Similarity  42.7%;  Pred. No. 1.85e-32;
  Matches    35;  Conservative   15;  Mismatches  25;  Indels   7;  Gaps   2;

Db    287 NINECASGPCQNGGTCVDGVNGFVCQCPPNYTGTYCE-----ISLDA--CSSMPCQNGAT 339
          | ::| :  |::|: ||| |||: | ||  ::| :||     : |::  |    |||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    340 CVNVGANYICECPPGFAGQNCE 361
          |: |  :  | |||||||  ||
Qy     61 CIVVQQEPTCRCPPGFAGPRCE 82


RESULT    8
ID   O42347      PRELIMINARY;      PRT;  1212 AA.
AC   O42347;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C-SERRATE-2 (FRAGMENT).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97184054.
RA   HAYASHI H., MOCHII M., KODAMA R., HAMADA Y., MIZUNO N., EGUCHI G.,
RA   TACHI C.;
RT   "Isolation of a novel chick homolog of Serrate and its coexpression
RT   with C-Notch-1 in chick development.";
RL   INT. J. DEV. BIOL. 40:1089-1096(1996).
DR   EMBL; D87558; D1022568; -.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   1212 AA;  134188 MW;  0ECF076C CRC32;

  Query Match           26.3%;  Score 212;  DB 13;  Length 1212;
  Best Local Similarity  39.3%;  Pred. No. 6.99e-29;
  Matches    33;  Conservative   19;  Mismatches  25;  Indels   7;  Gaps   4;

Db    468 ETNECESNPCQNGGRCKDLVNGFTCLCAQGFSGVFCE----MDI-D-F-CEPNPCQNGAK 520
          : ::| :: |::|::| | |||:||:|:|||||:|||    | : :   |:   |||||
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db    521 CYDLGGDYYCACPDDYDGKNCSHL 544
          |  :  :  | ||  : |  |  |
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKL 84


RESULT    9
ID   Q90819      PRELIMINARY;      PRT;  1193 AA.
AC   Q90819;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C-SERATE-1 PROTEIN (FRAGMENT).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
```

```
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 96175595.
RA   MYAT A., HENRIQUE D., ISH-HOROWICZ D., LEWIS J.;
RT   "A chick homologue of Serrate and its relationship with Notch and
RT   Delta homologues during central neurogenesis.";
RL   DEV. BIOL. 174:233-247(1996).
DR   EMBL; X95283; E224084; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 10.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   1193 AA;  131039 MW;  55E5FCD1 CRC32;

  Query Match           25.9%;  Score 209;  DB 13;  Length 1193;
  Best Local Similarity  37.8%;  Pred. No. 3.59e-28;
  Matches    31;  Conservative   18;  Mismatches  26;  Indels   7;  Gaps   2;

Db    463 NECASNPCMNGGHCQDEINGFQCLCPAGFSGNLC-Q------LDIDYCEPNPCQNGAQCF 515
          ::| :: | :|::| ||:||: |:|| |||| :| :      |:   |:   |||||||:
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    516 NLAMDYFCNCPEDYEGKNCSHL 537
          :  :  | ||  : |  |  |
Qy     63 VVQQEPTCRCPPGFAGPRCEKL 84


RESULT   10
ID   O42374      PRELIMINARY;      PRT;   752 AA.
AC   O42374;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH RECEPTOR PROTEIN (FRAGMENT).
GN   NOTCH6.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WESTIN J., LARDELLI M.;
RL   DEV. GENES EVOL. 207:51-63(1997).
DR   EMBL; Y10354; E293438; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 9.
DR   PROSITE; PS01186; EGF_2; 15.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 16.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
FT   NON_TER     752    752
SQ   SEQUENCE   752 AA;  82103 MW;  72E254FB CRC32;

  Query Match           25.8%;  Score 208;  DB 13;  Length 752;
  Best Local Similarity  38.0%;  Pred. No. 6.18e-28;
  Matches    30;  Conservative   15;  Mismatches  29;  Indels   5;  Gaps   3;

Db    462 GPRCKNGGQCVDGVGRYTCNCPPGFAGEHCEGDVNEC--RSGPC--YS-PGTIDCVPLIN 516
          | :|::|:|||| |  ||| || ||:|  ||       :::||  |      :|: :
Qy      7 GHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQQ 66

Db    517 SYQCRCRLGYTGQRCESMV 535
            |||  |::| ||| ::
Qy     67 EPTCRCPPGFAGPRCEKLI 85


RESULT   11
ID   Q25059      PRELIMINARY;      PRT;   406 AA.
AC   Q25059;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN III (FRAGMENT).
OS   HELIOCIDARIS ERYTHROGRAMMA (SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA; EUECHINOIDEA;
OC   ECHINACEA; ECHINOIDA; ECHINOMETRIDAE; HELIOCIDARIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BISGROVE B.W.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L33862; G499688; -.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 7.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   406 AA;  43475 MW;  45E6EE2C CRC32;

  Query Match           25.4%;  Score 205;  DB 5;  Length 406;
  Best Local Similarity  41.2%;  Pred. No. 3.15e-27;
  Matches    35;  Conservative   13;  Mismatches  30;  Indels   7;  Gaps   4;

Db     12 DGDDCNPNPCQNGAACIDQVNDYECICPPGFTGDNCE--TD-I--DV--CASAPCRNGGA 64
          : |||  : |::|| |:|:|| | |||| ||:|  ||     :  :   |    |:||:
Qy      1 NNDDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQ 60

Db     65 CVDGVNGYTCNCIPGFDGDNCENNI 89
          |:      || | ||| |  ||: |
Qy     61 CIVVQQEPTCRCPPGFAGPRCEKLI 85


RESULT   12
ID   Q19350      PRELIMINARY;      PRT;  1722 AA.
AC   Q19350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO EGF-LIKE REPEATS. NCBI GI: 1125776.
GN   F11C7.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A., VETTER J.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U42839; G1125776; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PFAM; PF00008; EGF; 24.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
```

```
SQ   SEQUENCE   1722 AA;  188383 MW;  CCFB86B8 CRC32;

  Query Match           25.2%;  Score 203;  DB 5;  Length 1722;
  Best Local Similarity  38.6%;  Pred. No. 9.30e-27;
  Matches   32;  Conservative   22;  Mismatches  21;  Indels   8;  Gaps   6;

Db    163 DEDECKENFCQNGADC-ENLKGSYECKCLKGFSGKYCEIQDKK---QCTS-D-Y-CHNNG 215
          ::|:|  : |::||:| ::::| | | |  |||| :||    :   | :: | | |:| :
Qy      1 NNDDCVGHKCRHGAQCVDEVNG-YTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGA 59

Db    216 QCISTGSDLSCKCSPGFDGAFCE 238
          |||    : :|:|:||| |: ||
Qy     60 QCIVVQQEPTCRCPPGFAGPRCE 82


RESULT   13
ID   P87357      PRELIMINARY;      PRT;   717 AA.
AC   P87357;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAD TRANSMEMBRANE PROTEIN PRECURSOR.
GN   DELTAD.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97346722.
RA   DORNSEIFER P., TAKKE C., CAMPOS-ORTEGA J.A.;
RT   "Overexpression of a zebrafish homologue of the Drosophila neurogenic
RT   gene Delta perturbs differentiation of primary neurons and somite
RT   development.";
RL   MECH. DEV. 63:159-171(1997).
DR   EMBL; Y11760; E307461; -.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 6.
KW   SIGNAL; TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    717       DELTAD TRANSMEMBRANE PROTEIN.
SQ   SEQUENCE   717 AA;  79061 MW;  5CC32ECA CRC32;

  Query Match           24.7%;  Score 199;  DB 13;  Length 717;
  Best Local Similarity  40.5%;  Pred. No. 8.07e-26;
  Matches   32;  Conservative   14;  Mismatches  26;  Indels   7;  Gaps   4;

Db    405 DHCSSNPCSNDAQCLDLVDSYLCQCPEGFTGTHCEDN---I--D-E-CATYPCQNGGTCQ 457
          | | :: | : |||:| |::| | ||:||:|  ||     :  :   |  | ||||: |
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    458 DGLSDYTCTCPPGYTGKNC 476
            : || ||||::|  |
Qy     63 VVQQEPTCRCPPGFAGPRC 81


RESULT   14
ID   O57462      PRELIMINARY;      PRT;   802 AA.
AC   O57462;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAA.
GN   DELTAA.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98165392.
RA   APPEL B., EISEN J.S.;
RT   "Regulation of neuronal specification in the zebrafish spinal cord by
RT   Delta function.";
RL   DEVELOPMENT 125:371-380(1998).
DR   EMBL; AF030031; G2809389; -.
DR   PROSITE; PS01186; EGF_2; 8.
KW   GLYCOPROTEIN.
SQ   SEQUENCE   802 AA;  88941 MW;  42F041BD CRC32;

  Query Match           24.7%;  Score 199;  DB 13;  Length 802;
  Best Local Similarity  38.1%;  Pred. No. 8.07e-26;
  Matches   32;  Conservative   16;  Mismatches  29;  Indels   7;  Gaps   2;

Db    447 DHCSSSPCSNGARCVDLVNSYLCQCPDGFTGMNCDRAGD----E---CSMYPCQNGGTCQ 499
          | | :  | :||:||| ||:| | ||:||:|: |:::      :   |  | ||||: |
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCI 62

Db    500 EGASGYMCTCPPGYTGRNCSSPVS 523
                | ||||::|  |   ::
Qy     63 VVQQEPTCRCPPGFAGPRCEKLIT 86

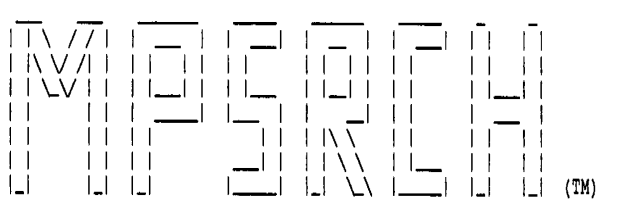

RESULT   15
ID   Q06007      PRELIMINARY;      PRT;   387 AA.
AC   Q06007;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH PROTEIN HOMOLOG 1 (MOTCH A PROTEIN) (FRAGMENT).
GN   NOTCH1 OR MOTCH A.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
DR   EMBL; X68278; G287988; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PFAM; PF00008; EGF; 6.
DR   PFAM; PF00066; notch; 3.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT.
FT   NON_TER      1      1
FT   NON_TER    387    387
SQ   SEQUENCE   387 AA;  41497 MW;  D1FD6C00 CRC32;

  Query Match           24.4%;  Score 197;  DB 11;  Length 387;
  Best Local Similarity  35.8%;  Pred. No. 2.37e-25;
  Matches   29;  Conservative   15;  Mismatches  36;  Indels   1;  Gaps   1;

Db     25 NECLSQPCQNGGTCIDLTNSYKCSCPRGTQGVHCEINVDDCHPPLDPASRSPKCFNNGTC 84
          ::|::: |::|: |:|  |:| | ||:|  |: ||          | :  | | : |
Qy      3 DDCVGHKCRHGAQCVDEVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQY-ECQNGAQC 61

Db     85 VDQVGGYTCTCPPGFVGERCE 105
          :      || ||||| | |||
Qy     62 IVVQQEPTCRCPPGFAGPRCE 82


Search completed: Fri May 28 09:37:15 1999
Job time : 33 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:39:01 1999; MasPar time 11.42 Seconds
452.494 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-14
Description: (1-243) from US09191647.pep
Perfect Score: 1905
Sequence: 1 ILDVASLRQAPGENGTSFHG.....SSFVDEVEKVVKCGCARCAS 243

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

tistics: Mean 31.391; Variance 131.102; scale 0.239

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1100	57.7	1534	30	W46966	Amino acid sequence o	2.21e-97
2	570	29.9	228	30	W46967	Amino acid sequence o	4.05e-44
3	499	26.2	183	30	W46968	Amino acid sequence o	4.14e-37
4	275	14.4	1872	36	W68510	Partial human Notch-3	1.76e-15
5	275	14.4	2321	36	W49698	Human Notch3 protein.	1.76e-15
6	263	13.8	612	28	W39256	Human partial mature	2.33e-14
7	263	13.8	737	28	W39257	Human membrane protei	2.33e-14
8	258	13.5	1036	25	W18351	Proliferation and dif	6.84e-14
9	258	13.5	1187	25	W18352	Proliferation and dif	6.84e-14
10	258	13.5	1208	28	W40827	Human Jagged protein.	6.84e-14
11	258	13.5	1218	29	W44301	Human serrate 1.	6.84e-14
12	258	13.5	1218	19	W05833	Human Serrate-1 (HJ1)	6.84e-14
13	256	13.4	1193	19	W05835	Chick Serrate.	1.05e-13
14	246	12.9	1218	25	W18354	Proliferation and dif	8.92e-13
15	245	12.9	1480	5	R25079	Drosophila SLIT prote	1.10e-12
16	238	12.5	1055	29	W44298	Human serrate 2 prote	4.90e-12
17	238	12.5	1212	29	W44299	Human serrate 2.	4.90e-12
18	238	12.5	1257	19	W05834	Human Serrate-2 (HJ2)	4.90e-12
19	234	12.3	473	17	R86869	Adhesive protein.	1.15e-11
20	230	12.1	833	6	R28960	Delta D11.	2.67e-11
21	216	11.3	727	21	W11719	C-Delta-1 polypeptide	5.12e-10
22	216	11.3	740	21	W00876	C-Delta-1 polypeptide	5.12e-10
23	213	11.2	722	21	W11720	M-Delta-1 polypeptide	9.60e-10
24	211	11.1	520	25	W18348	Proliferation and dif	1.46e-09
25	211	11.1	702	25	W18349	Proliferation and dif	1.46e-09
26	211	11.1	723	25	W18353	Proliferation and dif	1.46e-09
27	205	10.8	157	21	W11730	H-Delta-1 polypeptide	5.11e-09
28	205	10.8	660	21	W11725	H-Delta-1 polypeptide	5.11e-09
29	206	10.8	685	37	W80813	Nucleotide sequence o	4.15e-09
30	196	10.3	383	10	R56166	Neuroendocrine tumor	3.31e-08
31	196	10.3	1404	7	R38304	Sequence of a serrate	3.31e-08
32	191	10.0	1257	9	R46627	Neurocan core protein	9.30e-08
33	188	9.9	2409	3	R12609	Versican.	1.72e-07
34	179	9.4	385	10	R56167	Neuroendocrine tumor	1.09e-06
35	177	9.3	751	10	R53088	Human masking protein	1.64e-06
36	177	9.3	752	10	R53087	Human masking protein	1.64e-06
37	177	9.3	756	10	R53086	Human masking protein	1.64e-06
38	177	9.3	845	10	R53089	Human masking protein	1.64e-06
39	177	9.3	1355	3	R14584	TGF beta 1 binding pr	1.64e-06
40	177	9.3	1712	4	R22461	Masking protein high	1.64e-06
41	169	8.9	179	37	W75100	Human secreted protei	8.29e-06
42	166	8.7	379	5	R25565	Beta-IG-M1.	1.52e-05
43	163	8.6	77	6	R28962	ELR-11 and -12.	2.77e-05
44	162	8.5	559	25	W30844	Partial rat thrombomo	3.39e-05
45	162	8.5	577	25	W30845	Rat thrombomodulin.	3.39e-05

ALIGNMENTS

RESULT 1
ID W46966 standard; Protein; 1534 AA.
AC W46966;
DT 06-JUL-1998 (first entry)
DE Amino acid sequence of a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1534
FT /note= "mature protein"
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR N-PSDB; V16978.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 31-35; 45pp; Japanese.
CC The present sequence represents a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 1534 AA;

Query Match 57.7%; Score 1100; DB 30; Length 1534;
Best Local Similarity 55.9%; Pred. No. 2.21e-97;
Matches 128; Conservative 46; Mismatches 55; Indels 0; Gaps 0;

Db 1306 ngtgfhgcirnlyinnelqdftktqmkpgvvpgcepcrklyclhgicqpnatpgpmchce 1365
Qy 14 NGTSFHGCIRNLYINSELQDFRKMPMQTGILPGCEPCHKKVCAHGCCQPSSQSGFTCECE 73

Db 1366 agwvglhcdqpadgpchghkcvhgqcvpldalsyscqcqdgysgalcnqagalaepcrgl 1425

```
Qy     74 EGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQMI 133

Db   1426 qclhghcqasgtkgahcvcdpgfsgelceqesecrgdpvrdfhqvqrgyaicqttrplsw 1485
          | || |: ||   : | |::||:|: |::|  |||: :||: | |:||| ||||: :|:
Qy    134 KCKHGKCRLSGVGQPYCECNSGFTGDSCDREISCRGERIRDYYQKQQGYAACQTTKKVSR 193

Db   1486 vecrgscpgqgccqglrlkrrkftfecsdgtsfaeevekptkcgcalca 1534
           :||||:|:|  ||  || ||||::|||:||:|| :|||| |||| ||
Qy    194 LECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCARCA 242


RESULT    2
ID   W46967 standard; Protein; 228 AA.
AC   W46967;
DT   06-JUL-1998  (first entry)
DE   Amino acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody.
OS   Mus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   N-PSDB; V16966.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Page 35; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   228 AA;

  Query Match           29.9%;  Score 570;  DB 30;  Length 228;
  Best Local Similarity  57.3%;  Pred. No. 4.05e-44;
  Matches    67;  Conservative   20;  Mismatches  30;  Indels   0;  Gaps   0;

Db    110 ngtsfhgcirnlyinnelqdftktqmkpgvvpgcepcrklyclhgicqpnatpgpvchce 169
          |||||||||||||||:|||| |  |  |::||||||:|  | || |||:: :|  | ||
Qy     14 NGTSFHGCIRNLYINSELQDFRKMPMQTGILPGCEPCHKKVCAHGCCQPSSQSGFTCECE 73

Db    170 agwgglhcdqpvdgpchghkcvhgkcvpldalayscqcqdgysgalcnqvgavaepc 226
          || |  |||  : || |:|||| |:|::|::||| | :| :| ||::   : :||
Qy     74 EGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPC 130


ULT    3
   W46968 standard; Protein; 183 AA.
AC   W46968;
DT   06-JUL-1998  (first entry)
DE   Amino acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody.
OS   Rattus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   N-PSDB; V16967.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Pages 37-38; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   183 AA;

  Query Match           26.2%;  Score 499;  DB 30;  Length 183;
  Best Local Similarity  62.5%;  Pred. No. 4.14e-37;
  Matches    60;  Conservative   14;  Mismatches  22;  Indels   0;  Gaps   0;

Db     88 ngtsfhgcirnlyinnelqdftktqmkpgvvpgcepcrklyclhgicqpnatpgpvchce 147
          |||||||||||||||:||||| |  |  |::||||||:|  | || |||:: :|  | ||
Qy     14 NGTSFHGCIRNLYINSELQDFRKMPMQTGILPGCEPCHKKVCAHGCCQPSSQSGFTCECE 73

Db    148 agwgglhcdqpvdgpchghkcvhgkcvpldalaysc 183
          || |  |||  : || |:|||| |:|::|::|||
Qy     74 EGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSC 109


RESULT    4
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 328
FT                   /note= "encoded by NAN"
FT   Misc_difference 401
FT                   /note= "encoded by GNN"
FT   Misc_difference 403
FT                   /note= "encoded by GNC"
FT   Misc_difference 406
FT                   /note= "encoded by GNN"
FT   Misc_difference 409
FT                   /note= "encoded by NNT"
FT   Misc_difference 420
FT                   /note= "encoded by GNC"
FT   Misc_difference 706
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
```

```
SQ   Sequence   1872 AA;

  Query Match          14.4%;  Score 275;  DB 36;  Length 1872;
  Best Local Similarity  36.9%;  Pred. No. 1.76e-15;
  Matches   48;  Conservative   20;  Mismatches  50;  Indels  12;  Gaps  11;

Db     55 dpclsspcahgarcsvgpdgrflcscppgyqgrsc-rsdvdecrvgepcrhggtclntpg 113
          :||    ||||  |  :::: | | |  |  |  | :   | | :|: | || |||   :
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVHG-TCLPINA 104

Db    114 -sfrcqcpagytgplc-enpav--pcapspcrnggtcrqsg-dltydcaclpgfegqnce 168
           |: | |  |  | || |:  :  ||    |::|  || ||    | | | :|| |::|:
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCD 162

Db    169 vnvddcpghr 178
           ::  | | |
Qy    163 REIS-CRGER 171


RESULT    5
ID   W49698 standard; Protein; 2321 AA.
AC   W49698;
DT   21-DEC-1998  (first entry)
DE   Human Notch3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy.
OS   Homo sapiens.
PN   FR2751986-A1.
PD   06-FEB-1998.
PF   16-APR-1997; 004680.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133138/13.
DR   N-PSDB; V57001.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1.1-1.8; 45pp; French.
CC   This sequence represents the human Notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of
CC   the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC   and leukoencephalopathy (CADASIL) type.  Blocking expression of a
CC   mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC   gene or protein can be used to treat CADASIL or related disorders.
SQ   Sequence   2321 AA;

  Query Match          14.4%;  Score 275;  DB 36;  Length 2321;
  Best Local Similarity  36.9%;  Pred. No. 1.76e-15;
  Matches   48;  Conservative   20;  Mismatches  50;  Indels  12;  Gaps  11;

Db    121 dpclsspcahgarcsvgpdgrflcscppgyqgrsc-rsdvdecrvgepcrhggtclntpg 179
          :||    ||||  |  :::: | | |  |  |  | :   | | :|: | || |||   :
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVHG-TCLPINA 104

Db    180 -sfrcqcpagytgplc-enpav--pcapspcrnggtcrqsg-dltydcaclpgfegqnce 234
           |: | |  |  | || |:  :  ||    |::|  || ||    | | | :|| |::|:
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCD 162

Db    235 vnvddcpghr 244
           ::  | | |
Qy    163 REIS-CRGER 171


RESULT    6
ID   W39256 standard; protein; 612 AA.
AC   W39256;
DT   19-MAY-1998  (first entry)
DE   Human partial mature membrane protein.
KW   Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW   brain; nervous tissue; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Protein         1..612
FT                   /note= "partial mature protein"
PN   J10036395-A.
PD   10-FEB-1998.
PF   24-JUL-1996; 194467.
PR   24-JUL-1996; JP-194467.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-174912/16.
PT   New human membrane protein - specifically expressed in brain and
PT   nervous tissue; used in diagnosis of diseases specific to these
PT   tissues and cancer
PS   Claim 1; Column 18-19; 26pp; Japanese.
CC   W39256 represents the partial mature amino acid sequence of a novel
CC   membrane protein which contains epidermal growth factor (EGF) motifs.
CC   The new membrane protein is expressed specifically in brain and nervous
CC   tissue. The protein and DNA can be used in the diagnosis of brain and
CC   nerve system specific diseases and cancer.
SQ   Sequence   612 AA;

  Query Match          13.8%;  Score 263;  DB 28;  Length 612;
  Best Local Similarity  39.8%;  Pred. No. 2.33e-14;
  Matches   45;  Conservative   22;  Mismatches  35;  Indels  11;  Gaps  10;

Db    375 crngatci-sslsgftcqcpegyfgsaceekv-dpcasspcqnngtcy-vdgvhftcncs 431
          | :|  |  || |||||:| ||  |: |:::  ||| :: |  :|||  :::  ::|:|
Qy     55 CAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VHGTCLPINAFSYSCKCL 112

Db    432 pgftgptcaqlid-f--calspcahgtcr-s-vgtsykclcdpgyhglyceee 479
          |  |  | :  | |  | :  | || || | || :| | |::|: |  |: |
Qy    113 EGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDRE 164


RESULT    7
ID   W39257 standard; protein; 737 AA.
AC   W39257;
DT   19-MAY-1998  (first entry)
DE   Human membrane protein.
KW   Epidermal growth factor motif; EGF motif; membrane protein; disease;
KW   brain; nervous tissue; cancer; disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..26
FT                   /label= signal
FT   Protein         27..737
FT                   /label= membrane_protein
PN   J10036395-A.
PD   10-FEB-1998.
PF   24-JUL-1996; 194467.
PR   24-JUL-1996; JP-194467.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-174912/16.
DR   N-PSDB; V09641.
PT   New human membrane protein - specifically expressed in brain and
PT   nervous tissue; used in diagnosis of diseases specific to these
PT   tissues and cancer
PS   Claim 2; Pages 19-21; 26pp; Japanese.
CC   W39257 represents the amino acid sequence of a novel membrane protein
CC   which contains epidermal growth factor (EGF) motifs. The new membrane
CC   protein is expressed specifically in brain and nervous tissue. The
CC   protein and DNA can be used in the diagnosis of brain and nerve system
CC   specific diseases and cancer.
SQ   Sequence   737 AA;

  Query Match          13.8%;  Score 263;  DB 28;  Length 737;
  Best Local Similarity  39.8%;  Pred. No. 2.33e-14;
  Matches   45;  Conservative   22;  Mismatches  35;  Indels  11;  Gaps  10;

Db    401 crngatci-sslsgftcqcpegyfgsaceekv-dpcasspcqnngtcy-vdgvhftcncs 457
          | :|  |  || |||||:| ||  |: |:::  ||| :: |  :|||  :::  ::|:|
```

```
Qy     55 CAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VHGTCLPINAFSYSCKCL 112

Db    458 pgftgptcaqlid-f--calspcahgtcr-s-vgtsykclcdpgyhglyceee 505
Qy    113 EGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDRE 164


RESULT    8
ID   W18351 standard; protein; 1036 AA.
AC   W18351;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 5; Page 66-71; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1036 AA;

  Query Match           13.5%;  Score 258;  DB 25;  Length 1036;
  Best Local Similarity  37.0%;  Pred. No. 6.84e-14;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    566 nvcgphgkcksqsggkftcdcnkgftgtychenind-cesnpcrnggtcid-gvnsykci 623
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    624 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 678
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT    9
ID   W18352 standard; protein; 1187 AA.
AC   W18352;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
PR   30-NOV-1995; JP-311811.
PR   17-NOV-1995; JP-299611.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A,  Sakano S;
DR   WPI; 97-298110/27.
PT   Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT   proliferation and differentiation of undifferentiated human blood
PT   cells
PS   Claim 6; Page 71-76; 114pp; Japanese.
CC   The present sequence represents a polypeptide which suppresses
CC   proliferation and differentiation of undifferentiated cells such
CC   as neurons and blood cells. The polypeptide may be used for the
CC   prevention and control of disorders involving undifferentiated
CC   cells, such as leukaemia and malignant tumours, and improvement of
CC   blood formation, e.g. after immunosuppression.
SQ   Sequence   1187 AA;

  Query Match           13.5%;  Score 258;  DB 25;  Length 1187;
  Best Local Similarity  37.0%;  Pred. No. 6.84e-14;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    566 nvcgphgkcksqsggkftcdcnkgftgtychenind-cesnpcrnggtcid-gvnsykci 623
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    624 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 678
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT   10
ID   W40827 standard; Protein; 1208 AA.
AC   W40827;
DT   21-MAY-1998  (first entry)
DE   Human Jagged protein.
KW   Jagged; Notch; angiogenesis; endothelial cell; migration; human;
KW   wound repair; vulnerary; injury repair; signal transduction;
KW   motor neurone disease; amyotrophic lateral sclerosis; polymyelitis;
KW   diagnosis; therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..11
FT                   /label= Sig_peptide
FT   Domain          175..220
FT                   /note= "DSL (Delta, Serrate, Lag-2 and Apx-1)
FT                         domain"
FT   Region          224..852
FT                   /note= "EGF-like repeat region containing 16
FT                         EGF repeats"
FT   Misc_difference 526
FT                   /note= "encoded by ANC"
FT   Region          853..992
FT                   /note= "cysteine-rich region"
FT   Domain          1058..1083
FT                   /note= "transmembrane domain"
FT   Region          1084..1208
FT                   /note= "cytoplasmic region"
PN   WO9745143-A1.
PD   04-DEC-1997.
PF   30-MAY-1997; U09407.
PR   31-MAY-1996; US-018841.
PA   (NAAM-) NAT AMERICAN RED CROSS.
PA   (UYGE-) UNIV GENEVE.
PI   Maciag T, Montesano R, Pepper M, Wong MK, Zimrin AB;
DR   WPI; 98-032340/03.
DR   N-PSDB; V03674.
PT   New human Jagged protein - used to inhibit or promote angiogenesis
PT   and to control migration of endothelial cells in injured blood
PT   vessels
PS   Claim 2; Page 54-61; 81pp; English.
CC   This sequence comprises the human homologue of the rat Jagged
CC   protein.  Jagged is able to bind Notch protein and is involved in
CC   endothelial cell (EC) migration and differentiation.  The human
CC   Jagged amino acid sequence was deduced from a human endothelial
CC   cell cDNA (see V03674) induced by exposure to fibrin.  Jagged
CC   polypeptides can be expressed in host cell systems.  A method for
CC   treating or preventing disease by administering an agent that
CC   (ant)agonises, inhibits, prevents, enhances or stimulates function
CC   of the Notch or Jagged proteins is claimed, as well as a method for
CC   affecting differentiation of mesoderm, endoderm, ectoderm and/or
CC   neuroderm cells.  When Jagged is applied to a micro-diameter blood
CC   vessel from which ECs have been removed, damaged or reduced, it
CC   decrease migrations of EC to the site, but when delivered to a
CC   similar site on a large vessel it increases EC migration.  Jagged
CC   and its agonists are used to inhibit or prevent angiogenesis (where
```

```
CC   associated with solid tumours, rheumatoid arthritis, inflammation,
CC   or restenosis, particularly preventing angiogenesis from the vaso
CC   vasorum and promoting large vessel EC migration to repair the lumen
CC   of large vessels). Anti-Jagged and Jagged antagonists (e.g.
CC   antisense Jagged and Jagged mutants) are used to promote or enhance
CC   angiogenesis, particularly for wound and injury repair, e.g. where
CC   surgical, traumatic and/or caused by disease, e.g. diabetes-related
CC   (all claimed). Angiogenesis can be modulated in vitro or in vivo
CC   and expression of proteins by gene therapy is included. Modulation
CC   of the Notch-Jagged signalling pathway may also be involved in
CC   placental development and motor neurone diseases such as
CC   amyotrophic lateral sclerosis, poliomyelitis etc.
SQ   Sequence  1208 AA;

  Query Match          13.5%;  Score 258;  DB 28;  Length 1208;
  Best Local Similarity  37.0%;  Pred. No. 6.84e-14;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    587 nvcgphgkcksqsggkftcdcnkgftgtychenind-cesnpcrnggtcid-gvnsykci 644
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    645 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 699
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT  11
ID   W44301 standard; Protein; 1218 AA.
AC   W44301;
DT   19-JUN-1998  (first entry)
DE   Human serrate 1.
KW   Human; serrate 2; regulation; stem cell; differentiation; neoplasm;
KW   leukaemia; endothelial cell; tumour.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Signal
FT   Protein         32..1218
FT                   /label= Serrate-1
PN   WO9802458-A1.
PD   22-JAN-1998.
PF   11-JUL-1997; J02414.
PR   14-MAY-1997; JP-124063.
PR   16-JUL-1996; JP-186220.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
PI   Itoh A, Sakano S;
DR   WPI; 98-110528/10.
DR   N-PSDB; V15201.
PT   Human serrate-2 gene expression products  - used to regulate stem
PT   cell differentiation, useful in treating neoplasms, e.g. leukaemia
PS   Disclosure; Page 77-86; 103pp; Japanese.
CC   The present sequence represents human serrate 1, from the present
CC   invention which describes human serrate 2. The present invention also
CC   describes a method for the preparation of the polypeptides, and
CC   antibodies binding to the polypeptide and its fragments. The polypeptide
CC   and its fragments expressed by the serrate-2-gene can be used to inhibit
CC   stem (especially blood stem) cell differentiation and to inhibit
CC   endothelial cell growth. They may be incorporated in a cell culture
CC   media for culturing undifferentiated stem cells. They can also be used
CC   for treatment of neoplasms such as leukaemia. The antibodies can be used
CC   for the diagnosis of malignant tumours.
SQ   Sequence  1218 AA;

  Query Match          13.5%;  Score 258;  DB 29;  Length 1218;
  Best Local Similarity  37.0%;  Pred. No. 6.84e-14;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 nvcgphgkcksqsggkftcdcnkgftgtychenind-cesnpcrnggtcid-gvnsykci 654
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT  12
ID   W05833 standard; Protein; 1218 AA.
AC   W05833;
DT   28-JAN-1997  (first entry)
DE   Human Serrate-1 (HJ1).
KW   Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
KW   cell fate; central nervous system; cancer; tissue repair; therapy;
KW   diagnosis; antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1..1067
FT                   /label= Extracellular_domain
FT   peptide         14..29
FT                   /label= Sig_peptide
FT   domain          185..229
FT                   /label= DSL
FT                   /note= "region of homology with Drosophila Delta
FT                   and Serrate, predicted to mediate binding
FT                   with Notch"
FT   domain          234..896
FT                   /label= ELR
FT                   /note= "epidermal growth factor-like repeat domain"
FT   region          234..264
FT                   /label= ELR1
FT   region          265..299
FT                   /label= ELR2
FT   region          300..339
FT                   /label= ELR3
FT   region          340..377
FT                   /label= ELR4
FT   region          378..415
FT                   /label= ELR5
FT   region          416..453
FT                   /label= ELR6
FT   region          454..490
FT                   /label= ELR7
FT   region          491..528
FT                   /label= ELR8
FT   region          529..566
FT                   /label= ELR9
FT   region          567..598
FT                   /label= Partial_ELR
FT   region          599..632
FT                   /label= Partial_ELR
FT   region          633..670
FT                   /label= ELR10
FT   region          671..708
FT                   /label= ELR11
FT   region          709..747
FT                   /label= ELR12
FT   region          748..785
FT                   /label= ELR13
FT   region          786..823
FT                   /label= ELR14
FT   region          824..862
FT                   /label= ELR15
FT   region          863..879
FT                   /label= Partial_ELR
FT   region          880..896
FT                   /label= Partial_ELR
FT   domain          1068..1089
FT                   /label= Transmembrane_domain
FT   domain          1090..1218
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
```

```
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique DMP,  Ish-Horowicz D;
PI   Lewis JH, Mann RS,  Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40090.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Claim 4; Page 95-98; 161pp; English.
CC   Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC   for the zygotic neurogenic locus Notch, and are believed to play a
CC   major role in determining cell fates (differentiation) in the
CC   central nervous system.  Their amino acid sequences were deduced
CC   from cDNA clones (see also T40090-91) isolated from human foetal
CC   brain cDNA libraries.  The proteins, antibodies raised to them,
CC   and encoding nucleic acids can be used in the detection of
CC   Serrate sequences and in the treatment of disorders of cell fate
CC   or differentiation, partic. cancer, nervous system disorders
CC   and in tissue repair or regeneration.
SQ   Sequence   1218 AA;

  Query Match           13.5%;  Score 258;  DB 19;  Length 1218;
  Best Local Similarity  37.0%;  Pred. No. 6.84e-14;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 nvcgphgkcksqsggkftcdcnkgftgtychenind-cesnpcrnggtcid-gvnsykci 654
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT   13
ID   W05835 standard; Protein; 1193 AA.
AC   W05835;
DT   28-JAN-1997  (first entry)
DE   Chick Serrate.
KW   C-Serrate; Notch; cell differentiation; cell fate; tissue repair;
KW   central nervous system; cancer; therapy; diagnosis.
OS   Gallus sp.
FH   Key             Location/Qualifiers
FT   domain          1..1041
FT                   /label= Extracellular_domain
FT   peptide         1..5
FT                   /label= Sig_peptide
FT                   /note= "lacks the N-terminal portion owing to
FT                   truncation of the encoding cDNA clone"
FT   domain          158..203
FT                   /label= DSL
FT                   /note= "region of homology with Drosophila Delta
FT                   and Serrate, predicted to mediate binding
FT                   with Notch"
FT   domain          208..837
FT                   /label= ELR
FT                   /note= "epidermal growth factor-like repeat domain"
FT   region          208..238
FT                   /label= ELR1
FT   region          239..274
FT                   /label= ELR2
FT   region          275..313
FT                   /label= ELR3
FT   region          314..351
FT                   /label= ELR4
FT   region          352..390
FT                   /label= ELR5
FT   region          391..427
FT                   /label= ELR6
FT   region          428..464
FT                   /label= ELR7
FT   region          465..502
FT                   /label= ELR8
FT   region          503..540
FT                   /label= ELR9
FT   region          541..606
FT                   /label= ELR10
FT   region          607..644
FT                   /label= ELR11
FT   region          655..682
FT                   /label= ELR12
FT   region          683..721
FT                   /label= ELR13
FT   region          722..759
FT                   /label= ELR14
FT   region          760..797
FT                   /label= ELR15
FT   region          798..837
FT                   /label= ELR16
FT   region          854..911
FT                   /label= Cysteine-rich_region
FT   domain          1042..1066
FT                   /label= Transmembrane_domain
FT   domain          1067..1193
FT                   /label= Intracellular_domain
PN   WO9627610-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03172.
PR   07-MAR-1995; US-400159.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique DMP,  Ish-Horowicz D;
PI   Lewis JH, Mann RS,  Myat AM;
DR   WPI; 96-425379/42.
DR   N-PSDB; T40092.
PT   Vertebrate Serrate protein and related DNA - used to treat or
PT   prevent malignancies characterised by increased Notch activity.
PS   Disclosure; Page 112-115; 161pp; English.
CC   Chicken Serrate (W05835), or C-Serrate, is a ligand for the zygotic
CC   neurogenic locus Notch and is believed to play a major role in
CC   determining cell fates in the central nervous system.  Its amino
CC   acid sequence was deduced from a cDNA clone (T40092) obtd. from an
CC   optic explant cDNA library.  C-Serrate is expressed in the central
CC   nervous system, cranial placodes, nephric mesoderm, vascular
CC   system, and limb bud mesenchyme.
SQ   Sequence   1193 AA;

  Query Match           13.4%;  Score 256;  DB 19;  Length 1193;
  Best Local Similarity  37.0%;  Pred. No. 1.05e-13;
  Matches   44;  Conservative   21;  Mismatches  44;  Indels  10;  Gaps   9;

Db    571 nvcgphgkcksqaggkftcecnkgftgtychenind-cesnpcknggtcid-gvnsykci 628
          :||: || | : : : |||||: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    629 csdgwegtyc--etni-ndcsknpchnggtcr-dlvndffceckngwkgktchsrdsqc 683
          | :|  |  |  | :: | | :  | :|  ||   | : :|||::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT   14
ID   W18354 standard; protein; 1218 AA.
AC   W18354;
DT   11-FEB-1998  (first entry)
DE   Proliferation and differentiation suppression polypeptide.
KW   Proliferation; differentiation; suppression; human; delta-1;
KW   serrate-1; blood cell; neuron; leukaemia; malignant tumour;
KW   immunosuppression.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Signal
FT   Protein         32..1218
FT                   /label= Differentiation_suppression_protein
PN   WO9719172-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; J03356.
```

PR 30-NOV-1995; JP-311811.
PR 17-NOV-1995; JP-299611.
PA (ASAH) ASAHI KASEI KOGYO KK.
PI Itoh A, Sakano S;
DR WPI; 97-298110/27.
DR N-PSDB; T70175.
PT Peptide(s) encoded by human genes delta-1 and serrate-1 - suppress
PT proliferation and differentiation of undifferentiated human blood
PT cells
PS Claim 15; Page 83-91; 114pp; Japanese.
CC The present sequence represents a polypeptide which suppresses
CC proliferation and differentiation of undifferentiated cells such
CC as neurons and blood cells. The polypeptide may be used for the
CC prevention and control of disorders involving undifferentiated
CC cells, such as leukaemia and malignant tumours, and improvement of
CC blood formation, e.g. after immunosuppression.
Sequence 1218 AA;

Query Match 12.9%; Score 246; DB 25; Length 1218;
Best Local Similarity 36.1%; Pred. No. 8.92e-13;
Matches 43; Conservative 21; Mismatches 45; Indels 10; Gaps 9;

Db 597 nvcgphgkcksqsggkftcdcnkgftgtycheninds-esnpcrnggtcid-gvnsykci 654
Qy 53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db 655 csdgwegayc--etni-ndcsqnpchnggtcr-dlvndfycdckngwkgktchsrdsqc 709
Qy 111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167

RESULT 15
ID R25079 standard; Protein; 1480 AA.
AC R25079;
DT 05-JAN-1993 (first entry)
DE Drosophila SLIT protein involved in axon pathway development.
KW Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW embryonic CNS; leucine-rich repeat; Flank-LRR-Flank;
KW midline glial cells; axonogenesis; cell-cell interaction; ss.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT peptide 1..36
FT /label= signal
FT domain 73..294
FT /label= Flank_LRR_Flank_1
FT /note= "mediates adhesive events"
FT domain 295..518
FT /label= Flank-LRR-Flank_2
FT /note= "mediates adhesive events"
FT domain 519..714
FT /label= Flank_LRR_Flank_3
FT /note= "mediates adhesive events"
FT domain 715..910
FT /label= Flank_LRR_Flank_4
FT /note= "mediates adhesive events"
FT region 911..1150
FT /label= Tandem_EGF_like_repeats
FT /note= "involved in protein-protein interactions"
FT region 1353..1393
FT /label= 7th_EGF_like_repeat
FT /note= "involved in receptor-ligand interactions"
FT region 1394..1404
FT /label= alternative_splice_segment
FT /note= "developmentally regulated"
FT region 1405..1480
FT /label= C-terminal_region
PN WO9210518-A.
PD 25-JUN-1992.
PF 27-NOV-1991; U09055.
PR 07-DEC-1990; US-624135.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Rothberg JM;
DR WPI; 92-234590/28.
DR N-PSDB; Q25811.
PT SLIT protein and sequence elements for treating
PT neuro-degenerative disease - useful for Alzheimer's disease,
PT nerve damage and Parkinson's disease, for diagnosis of cancer
PS Claim 1; Page 84-89; 122pp; English.
CC The SLIT protein is necessary for normal development of the midline
CC of the CNS, partic. the midline glial cells, and for the
CC concomitant formation of the commisural axon pathways. The process
CC is dependent on the level of SLIT protein expression. It appears
CC that SLIT protein is excreted by the midline glial cells where it
CC is synthesised and is eventually associated with the surfaces of
CC axons that traverse tham. The SLIT protein is tightly localised to
CC the muscle attachment sites and to the sites of contact between
CC adjacent pairs of cardioblasts as they coalesce to form the lumen
CC of the larval heart. The SLIT protein defines a new set of
CC molecules (TAGONS) which play a key role in axon outgrowth and
CC pathfinding. SLIT can be used as a nerve regenerative in
CC neurodegenerative diseases such as Alzheimer's Disease, spinal cord
CC injuries, brain injuries, crushed (optic) nerve, amytrophic lateral
CC sclerosis, diabetes-caused nerve damage, Parkinson's Disease,
CC strokes, epilepsy, multiple sclerosis, paraplegia retinal
CC degeneration and facial nerve damage. The 4 "Flank-Leucine-rich
CC region-Flank" domains, the C-terminal region and part of the
CC alternative splice segment (i.e. GEGSTEPFTVT) are all individually
CC claimed as are molecules comprising at least 1 FLank-LRR-Flank domain
CC and at least 1 EGF-like repeat element from the SLIT protein.
CC See also R29102.
SQ Sequence 1480 AA;

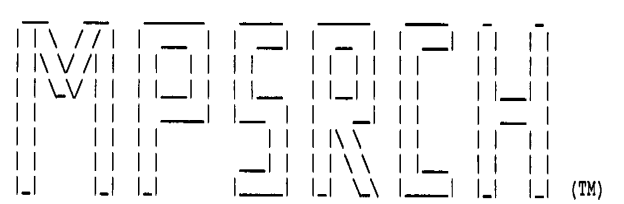
Query Match 12.9%; Score 245; DB 5; Length 1480;
Best Local Similarity 31.9%; Pred. No. 1.10e-12;
Matches 44; Conservative 30; Mismatches 52; Indels 12; Gaps 11;

Db 900 vrndilakcnacfeqpcqnqaqcvalpqreyqclcqpgyhgkhcefmi-dacygnpcrnn 958
Qy 39 MQTGILPGCEPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VH 96

Db 959 atctvleegrfrcgcapgytgarcetnid--d-clgeikcqnnatc-idgv-esykcecq 1013
Qy 97 GTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQ-MIKC-KHGKCRLSGVGQPY-CECN 153

Db 1014 pgfsgefcdtkiqfcspe 1031
Qy 154 SGFTGDSCDREIS-CRGE 170

Search completed: Fri May 28 09:40:06 1999
Job time : 65 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

?srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:40:24 1999; MasPar time 12.37 Seconds
786.870 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-14
Description: (1-243) from US09191647.pep
Perfect Score: 1905
Sequence: 1 ILDVASLRQAPGENGTSFHG.....SSFVDEVEKVVKCGCARCAS 243

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 40.820; Variance 67.040; scale 0.609

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	292	15.3	1429	2	S06434	homeotic protein lin-	1.04e-42
2	290	15.2	1203	2	A49175	Motch B protein - mou	2.98e-42
3	285	15.0	2471	2	A49128	cell-fate determining	4.14e-41
4	274	14.4	2318	2	S45306	notch 3 protein - mou	1.32e-38
5	275	14.4	2321	2	S78549	notch3 protein - huma	7.83e-39
6	269	14.1	1220	2	A56136	jagged protein precur	1.79e-37
7	264	13.9	1469	2	B36665	slit protein 2 precur	2.41e-36
8	262	13.8	2524	2	A35844	Xotch protein - Afric	6.82e-36
9	260	13.6	2531	2	S18188	notch protein homolog	1.92e-35
10	260	13.6	2531	2	A46019	gene Notch-1 protein	1.92e-35
11	254	13.3	530	2	A31640	epidermal growth fact	4.27e-34
12	254	13.3	1480	2	A36665	slit protein 1 precur	4.27e-34
13	251	13.2	2437	2	S42612	transmembrane protein	2.00e-33
14	245	12.9	2139	2	A35672	crumbs protein - frui	4.36e-32
15	246	12.9	2555	2	A40043	notch protein homolog	2.61e-32
16	244	12.8	2703	2	A24420	notch protein - fruit	7.27e-32
17	242	12.7	570	2	A48836	fibropellin C precurs	2.02e-31
18	239	12.5	1295	2	A32901	glp1 protein precurso	9.35e-31
19	234	12.3	473	2	A56175	adhesive plaque prote	1.19e-29
20	234	12.3	832	2	A31246	neurogenic protein De	1.19e-29
21	234	12.3	880	2	S00670	gene Delta protein pr	1.19e-29
22	233	12.2	833	2	S19087	gene Delta protein pr	1.98e-29
23	230	12.1	1064	2	A40136	fibropellin Ia - sea	9.05e-29
24	231	12.1	4391	2	A38096	perlecan precursor -	5.45e-29
25	219	11.5	861	2	A48825	Notch homolog Motch p	2.30e-26
26	216	11.3	728	2	I50719	C-Delta-1 - chicken	1.03e-25
27	213	11.2	387	2	B49175	Motch A protein - mou	4.61e-25
28	213	11.2	722	2	I48324	DELTA-like 1 - mouse	4.61e-25
29	211	11.1	293	2	B26637	neurogenic repetitive	1.25e-24
30	212	11.1	5147	1	IJFFTM	cadherin-related tumo	7.58e-25
31	208	10.9	3707	2	S18252	heparan sulfate prote	5.52e-24
32	203	10.7	200	2	A26637	neurogenic repetitive	6.52e-23
33	203	10.7	1404	2	A36666	serrate protein precu	6.52e-23
34	203	10.7	1408	2	S16148	gene serrate protein	6.52e-23
35	196	10.3	259	2	S48713	fetal antigen 1 - hum	2.02e-21
36	197	10.3	383	2	B45484	delta-like dlk homeot	1.24e-21
37	197	10.3	383	2	S53716	homeotic protein dlk	1.24e-21
38	197	10.3	1268	2	S52781	neurocan - mouse	1.24e-21
39	195	10.2	1257	2	S28764	neurocan - rat	3.29e-21
40	189	9.9	260	2	A44549	fetal antigen 1 homeo	6.08e-20
41	188	9.9	2409	2	A60979	versican precursor -	9.86e-20
42	185	9.7	1959	1	AGRT	agrin - rat	4.19e-19
43	185	9.7	2397	2	A55535	versican precursor -	4.19e-19
44	183	9.6	862	2	S43922	versican - pig-tailed	1.10e-18
45	179	9.4	385	2	S53718	homeotic protein dlk	7.44e-18

ALIGNMENTS

RESULT 1
ENTRY S06434 #type complete
TITLE homeotic protein lin-12 precursor - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 12-Sep-1997
ACCESSIONS S06434; A24769
REFERENCE S06434
#authors Yochem, J.; Weston, K.; Greenwald, I.
#journal Nature (1988) 335:547-550
#title The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with overall similarity to Drosophila Notch.
#cross-references MUID:88334747
#accession S06434
##molecule_type DNA
##residues 1-1429 ##label YOC
##cross-references EMBL:M12069; NID:g156357; PID:g156358
REFERENCE A24769
#authors Greenwald, I.
#journal Cell (1985) 43:583-590
#cross-references MUID:86079540
#accession A24769
##molecule_type DNA
##residues 173-712 ##label GRE
GENETICS
#introns 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3
CLASSIFICATION #superfamily unassigned ankyrin repeat proteins; ankyrin repeat homology
KEYWORDS glycoprotein; transmembrane protein
FEATURE
909-931 #domain transmembrane #status predicted #label TMM\
1093-1125 #domain ankyrin repeat homology #label AN1\
1206-1238 #domain ankyrin repeat homology #label AN2\
1240-1272 #domain ankyrin repeat homology #label AN3
SUMMARY #length 1429 #molecular-weight 157114 #checksum 4196

Query Match 15.3%; Score 292; DB 2; Length 1429;
Best Local Similarity 40.5%; Pred. No. 1.04e-42;
Matches 47; Conservative 24; Mismatches 35; Indels 10; Gaps 9;

```
Db    333 ICNHGTCIDSPLSEKAFECQCEPGYEGILCEQDKNE-CLSENMCLNNGTCVNLPG-SFRC 390
          :| || | : | |: :| |:|| |  | ||:|  |: ||: |:|:: |||: : : |: |
Qy     54 VCAHGCC-Q-PSSQSGFTCECEEGWMGPLCDQRTNDPCLG-NKCVH-GTCLPINAFSYSC 109

Db    391 DCARGFGGKWCDEP--L-NMCQDFHCENDGTCMHTSDHSPVCQCKNGFIGKRCEKE 443
          |  | ||  |||   | | || : | : | |  ::   | |:|::|| |  |::|
```

```
Qy    110 KCLEGHGGVLCDEEEDLFNPCQMIKCKH-GKCRLSGVGQPYCECNSGFTGDSCDRE 164


RESULT    2
ENTRY           A49175     #type fragment
TITLE           Motch B protein - mouse (fragment)
ALTERNATE_NAMES Notch homolog
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change
                  14-Aug-1998
ACCESSIONS      A49175; PH1570; S32113
REFERENCE       A49175
   #authors     Lardelli, M.; Lendahl, U.
   #journal     Exp. Cell Res. (1993) 204:364-372
   #title       Motch A and Motch B--two mouse Notch homologues coexpressed
                  in a wide variety of tissues.
   #cross-references MUID:93178563
   #accession   A49175
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues     1-1203 ##label LAR
      ##cross-references EMBL:X68279; NID:g287989; PID:g287990
      ##experimental_source embryo
      ##note         sequence extracted from NCBI backbone (NCBIP:126158)
COMMENT    This protein has many EGF repeats and lin-12/Notch repeats.
COMMENT    This protein is one of the neurogenic proteins controlling the
             decision between ectodermaland neural fate for cells in the early
             embryo.
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   560-591          #domain EGF homology #label EGF
SUMMARY         #length 1203  #checksum 910

  Query Match           15.2%;  Score 290;  DB 2;  Length 1203;
  Best Local Similarity  33.3%;  Pred. No. 2.98e-42;
  Matches   43;  Conservative   34;  Mismatches  41;  Indels  11;  Gaps  10;

Db    255 DNCDPDPCHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 311
          : |    | || || : :| :|| |: | ||::| :: :: | :: |:: | |: :   :
Qy     48 EPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 106

Db    312 YQCNCQPGTSGLNC--EIN-FDDCASNPCMHGVC-VDGINR-YSCVCSPGFTGQRCNIDI 366
          | |:|  | :|: |  | : |: |    |:|| | : |: : | | |::|||| |: :|
Qy    107 YSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDREI 165

Db    367 DECASNPCR 375
            | ::  |
Qy    166 S-CRGERIR 173


RESULT    3
ENTRY           A49128     #type complete
TITLE           cell-fate determining gene Notch2 protein - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
                  14-Aug-1998
ACCESSIONS      A49128
REFERENCE       A49128
   #authors     Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal     Development (1992) 116:931-941
   #title       Notch2: a second mammalian Notch gene.
   #cross-references MUID:93202015
   #accession   A49128
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-2471 ##label WEI
      ##experimental_source Schwann cell
      ##note         sequence extracted from NCBI backbone (NCBIP:127811)
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   1029-1060        #domain EGF homology #label EGF\
   1876-1908        #domain ankyrin repeat homology #label AN1\
   1909-1941        #domain ankyrin repeat homology #label AN2\
   1943-1975        #domain ankyrin repeat homology #label AN3\
   1976-2008        #domain ankyrin repeat homology #label AN4\
   2009-2041        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2471  #molecular-weight 265367  #checksum 5929

  Query Match           15.0%;  Score 285;  DB 2;  Length 2471;
  Best Local Similarity  33.3%;  Pred. No. 4.14e-41;
  Matches   43;  Conservative   33;  Mismatches  42;  Indels  11;  Gaps  10;

Db    572 DNCDPDPCHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 628
          : |    | || || : :| :|| |: | ||::| :: :: | :: |:: | |: :   :
Qy     48 EPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 106

Db    629 YQCNCQPGTSGLNC--EIN-FDDCASNPCLHGAC-VDGINR-YSCVCSPGFTGQRCNIDI 683
          | |:|  | :|: |  | : |: |    | || | : |: : | | |::|||| |: :|
Qy    107 YSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDREI 165

Db    684 DECASNPCR 692
            | ::  |
Qy    166 S-CRGERIR 173


RESULT    4
ENTRY           S45306     #type complete
TITLE           notch 3 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S45306
REFERENCE       S45306
   #authors     Lardelli, M.; Dahlstrand, J.; Lendahl, U.
   #journal     Mech. Dev. (1994) 46:123-136
   #title       The novel Notch homologue mouse Notch 3 lacks specific
                  epidermal growth factor-repeats and is expressed in
                  proliferating neuroepithelium.
   #cross-references MUID:95001556
   #accession   S45306
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2318 ##label LAR
      ##cross-references EMBL:X74760; NID:g483580; PID:g483581
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1839-1871        #domain ankyrin repeat homology #label AN1\
   1872-1904        #domain ankyrin repeat homology #label AN2\
   1906-1938        #domain ankyrin repeat homology #label AN3\
   1939-1971        #domain ankyrin repeat homology #label AN4\
   1972-2004        #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2318  #molecular-weight 244245  #checksum 9358

  Query Match           14.4%;  Score 274;  DB 2;  Length 2318;
  Best Local Similarity  36.2%;  Pred. No. 1.32e-38;
  Matches   47;  Conservative   21;  Mismatches  50;  Indels  12;  Gaps  11;

Db    122 DPCVSRPCVHGAPCSVGPDGRFACACPPGYQGQSC-QSDIDECRSGTTCRHGGTCLNTPG 180
          :||  : | ||  |  :::: |:| |  |  |  | | |   | |  |  | || |||   :
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCL-GNKCVHG-TCLPINA 104

Db    181 -SFRCQCPLGYTGLLC--ENPVV-PCAPSPCRNGGTCRQSS-DVTYDCACLPGFEGQNCE 235
           |: | |  |  |:||  |: :  ||    |::|  || |:    | | | :|| |::|:
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCD 162

Db    236 VNVDDCPGHR 245
            ::  | | |
Qy    163 REIS-CRGER 171


RESULT    5
ENTRY           S78549     #type complete
TITLE           notch3 protein - human
```

```
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                  17-Mar-1999
ACCESSIONS      S78549; S71825
REFERENCE       S78549
  #authors      Joutel, A.; Tournier-Lasserve, E.
  #submission   submitted to the EMBL Data Library, April 1997
  #accession    S78549
    ##molecule_type mRNA
    ##residues     1-2321 ##label JOU1
    ##cross-references EMBL:U97669; NID:g2668591; PID:g2668592
REFERENCE       S71825
  #authors      Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat,
                  H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.;
                  Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.;
                  Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.;
                  Bousser, M.G.; Tournier-Lasserve, E.
  #journal      Nature (1996) 383:707-710
  #title        Notch3 mutations in CADASIL, a hereditary adult-onset
                  condition causing stroke and dementia.
  #cross-references MUID:97032728
  #accession    S71825
    ##status       nucleic acid sequence not shown
    ##molecule_type DNA
    ##residues     67-113;138-194;268-333,'G',335-346;536-613;716-765;
                    1240-1279;1815-1888 ##label JOU2
    ##cross-references EMBL:U97669
GENETICS
  #gene         notch3
  #map_position 19p13.1
FUNCTION
  #description  may be involved in pathogenesis of CADASIL, causing a type of
                  stroke and dementia
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
KEYWORDS        tandem repeat; transmembrane protein
FEATURE
  318-349           #domain EGF homology #label EGF\
  1838-1870         #domain ankyrin repeat homology #label AN1\
  1871-1903         #domain ankyrin repeat homology #label AN2\
  1905-1937         #domain ankyrin repeat homology #label AN3\
  1938-1970         #domain ankyrin repeat homology #label AN4\
  1971-2003         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2321  #molecular-weight 243657  #checksum 3337

Query Match           14.4%;  Score 275;  DB 2;  Length 2321;
Best Local Similarity  36.9%;  Pred. No. 7.83e-39;
Matches   48;  Conservative   20;  Mismatches  50;  Indels  12;  Gaps  11;

Db    121 DPCLSSPCAHGARCSVGPDGRFLCSCPPGYQGRSC-RSDVDECRVGEPCRHGGTCLNTPG 179
          :||    |||| |  :::: | | |  |  |  | :   | | :|: | || |||   :
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVHG-TCLPINA 104

Db    180 -SFRCQCPAGYTGPLC-ENPAV--PCAPSPCRNGGTCRQSG-DLTYDCACLPGFEGQNCE 234
           |: | |  | | || |:  :  ||   |::|  || ||    | | | :|| |::|:
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCD 162

Db    235 VNVDDCPGHR 244
          ::  | | |
Qy    163 REIS-CRGER 171


RESULT    6
ENTRY           A56136     #type complete
TITLE           jagged protein precursor - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change
                  11-Aug-1995
ACCESSIONS      A56136
REFERENCE       A56136
  #authors      Lindsell, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.
  #journal      Cell (1995) 80:909-917
  #title        Jagged: a mammalian ligand that activates Notch1.
  #cross-references MUID:95211842
  #accession    A56136
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1220 ##label LIN
    ##cross-references GB:L38483
SUMMARY         #length 1220  #molecular-weight 134528  #checksum 2746

  Query Match           14.1%;  Score 269;  DB 2;  Length 1220;
  Best Local Similarity  37.8%;  Pred. No. 1.79e-37;
  Matches   45;  Conservative   19;  Mismatches  45;  Indels  10;  Gaps   9;

Db    598 NVCGPHGKCKSESGGKFTCDCNKGFTGTYCHENIND-CEGNPCTNGGTCID-GVNSYKCI 655
          :||: || | : | : |||:|: |  |  | :  || | || | :| ||:     || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    656 CSDGWEGAHC--ENNI-NDCSQNPCHYGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 710
          | :|  |  |  |::: | |    |  |  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT    7
ENTRY           B36665     #type complete
TITLE           slit protein 2 precursor - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                  16-Dec-1998
ACCESSIONS      B36665
REFERENCE       A36665
  #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                  Artavanis-Tsakonas, S.
  #journal      Genes Dev. (1990) 4:2169-2187
  #title        slit: an extracellular protein necessary for development of
                  midline glia and commissural axon pathways contains both
                  EGF and LRR domains.
  #cross-references MUID:91099665
  #accession    B36665
    ##status       preliminary
    ##molecule_type mRNA
    ##residues     1-1469 ##label ROT
    ##cross-references GB:X53959
GENETICS
  #gene         FlyBase:sli
    ##cross-references FlyBase:FBgn0003425
CLASSIFICATION  #superfamily proteoglycan amino-terminal homology; EGF
                  homology; leucine-rich alpha-2-glycoprotein repeat
                  homology; proteoglycan carboxyl-terminal homology
FEATURE
  66-91             #domain proteoglycan amino-terminal homology #label
                      PAH1\
  101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR1\
  125-148           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR2\
  149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR3\
  173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR4\
  197-220           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR5\
  228-272           #domain proteoglycan carboxyl-terminal homology #label
                      PCS1\
  288-313           #domain proteoglycan amino-terminal homology #label
                      PAH2\
  323-346           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR6\
  347-370           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR7\
  371-394           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR8\
  395-418           #domain leucine-rich alpha-2-glycoprotein repeat
                      homology #label LRR9\
```

```
  419-442          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR10\
  450-494          #domain proteoglycan carboxyl-terminal homology #label
                     PCS2\
  512-537          #domain proteoglycan amino-terminal homology #label
                     PAH3\
  547-571          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR11\
  572-595          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR12\
  596-619          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR13\
  620-643          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR14\
  651-695          #domain proteoglycan carboxyl-terminal homology #label
                     PCS3\
  708-733          #domain proteoglycan amino-terminal homology #label
                     PAH4\
  743-766          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR15\
  767-790          #domain leucine-rich alpha-2-glycoprotein repeat
                     homology #label LR16\
  846-890          #domain proteoglycan carboxyl-terminal homology #label
                     PCS4\
  1028-1061        #domain EGF homology #label EGF
SUMMARY          #length 1469  #molecular-weight 164695  #checksum 8361

  Query Match           13.9%;  Score 264;  DB 2;  Length 1469;
  Best Local Similarity  34.1%;  Pred. No. 2.41e-36;
  Matches   42;  Conservative   27;  Mismatches  44;  Indels  10;  Gaps   9;

Db   1351 EEPVDPCLENKCRRGSRCVPNSNARDGYQCKCKHGQRGRYCDQAASTCRKEQVREYY-TE 1409
           |:  :||   ||::| :|   |   : | | |: |  |  ||:  | || |::|:||  :
Qy    124 EDLFNPCQMIKCKHG-KC-RLSGVGQPY-CECNSGFTGDSCDREIS-CRGERIRDYYQKQ 179

Db   1410 NDCRSRQPLK-YAK--CVGGC-GNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGCT 1465
             : |  |  ::  | ||| | |||::   :|||    |::   :: ::: | ||||:
Qy    180 QGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCA 239

Db   1466 KKC 1468
           : |
Qy    240 R-C 241


RESULT    8
ENTRY            A35844     #type complete
TITLE            Xotch protein - African clawed frog
ORGANISM         #formal_name Xenopus laevis #common_name African clawed frog
DATE             12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change
                   14-Aug-1998
ACCESSIONS       A35844
REFERENCE        A35844
   #authors      Coffman, C.; Harris, W.; Kintner, C.
   #journal      Science (1990) 249:1438-1441
   #title        Xotch, the Xenopus homolog of Drosophila notch.
   #cross-references MUID:90385285
   #accession    A35844
      ##status        preliminary; nucleic acid sequence not shown; not
                       compared with conceptual translation
      ##molecule_type mRNA
      ##residues      1-2524 ##label COF
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology; EGF homology
KEYWORDS         transmembrane protein
FEATURE
   222-254          #domain EGF homology #label EGF\
   1924-1956        #domain ankyrin repeat homology #label AN1\
   1957-1989        #domain ankyrin repeat homology #label AN2\
   1991-2023        #domain ankyrin repeat homology #label AN3\
   2024-2056        #domain ankyrin repeat homology #label AN4\
   2057-2089        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2524  #molecular-weight 274931  #checksum 9441

  Query Match           13.8%;  Score 262;  DB 2;  Length 2524;
  Best Local Similarity  31.9%;  Pred. No. 6.82e-36;
  Matches   43;  Conservative   31;  Mismatches  49;  Indels  12;  Gaps  11;

Db    714 LSEVNECNSNPCIHGACHDGV-NGYKCDCEAGWSGSNCDINNNE-CESNPCMNGGTCKDM 771
          |:  : |: : | || |: :  :|: |:|| || |: ||   |: | :| |::| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TG-AYICTCKAGFSGPNC--QTNI-NECSSNPCLNHGTC-IDDVA-GYKCNCMLPYTGAI 825
           : :| | |  | :|  |  : :: | |    | :|| | :  |:  | |:|   :||
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CEAVLAPCAGSPCKN 840
          |:  :: | |   ::
Qy    161 CDREIS-CRGERIRD 174


RESULT    9
ENTRY            S18188     #type complete
TITLE            notch protein homolog - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
                   12-Feb-1999
ACCESSIONS       S18188
REFERENCE        S18188
   #authors      Weinmaster, G.; Roberts, V.J.; Lemke, G.
   #journal      Development (1991) 113:199-205
   #title        A homolog of Drosophila Notch expressed during mammalian
                   development.
   #cross-references MUID:92111383
   #accession    S18188
      ##molecule_type mRNA
      ##residues      1-2531 ##label WEI
      ##cross-references EMBL:X57405; NID:g57634; PID:g57635
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                   repeat homology
FEATURE
   1917-1949        #domain ankyrin repeat homology #label AN1\
   1950-1982        #domain ankyrin repeat homology #label AN2\
   1984-2016        #domain ankyrin repeat homology #label AN3\
   2017-2049        #domain ankyrin repeat homology #label AN4\
   2050-2082        #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2531  #molecular-weight 270907  #checksum 2705

  Query Match           13.6%;  Score 260;  DB 2;  Length 2531;
  Best Local Similarity  33.6%;  Pred. No. 1.92e-35;
  Matches   41;  Conservative   27;  Mismatches  43;  Indels  11;  Gaps  10;

Db    714 LSEVNECNSNPCIHGACR-DGLNGYKCDCAPGWSGTNCDINNNE-CESNPCVNGGTCKDM 771
          |:  : |: : | || |:  : :|: |:|  || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TS-GYVCTCREGFSGPNC--QTNI-NECASNPCLNQGTC-IDDVA-GYKCNCPLPYTGAT 825
           : :| | | || :|  |  : :: | |    | ::| | :  |:  | |:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CE 827
          |:
Qy    161 CD 162


RESULT   10
ENTRY            A46019     #type complete
TITLE            gene Notch-1 protein - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                   14-Aug-1998
ACCESSIONS       A46019
REFERENCE        A46019
   #authors      del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins,
                   N.A.; Copeland, N.G.; Gridley, T.
   #journal      Genomics (1993) 15:259-264
   #title        Cloning, analysis, and chromosomal localization of Notch-1, a
```

```
                mouse homolog of Drosophila Notch.
   #cross-references MUID:93194170
   #accession    A46019
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues      1-2531 ##label DEL
      ##cross-references GB:Z11886; GB:S47228; NID:g288502; PID:g288503
      ##note          sequence extracted from NCBI backbone (NCBIP:127318)
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   757-788           #domain EGF homology #label EGF\
   1917-1948         #domain ankyrin repeat homology #label AN1\
   1949-1981         #domain ankyrin repeat homology #label AN2\
   1983-2015         #domain ankyrin repeat homology #label AN3\
   2016-2048         #domain ankyrin repeat homology #label AN4\
   2049-2081         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2531  #molecular-weight 271312  #checksum 6611

  Query Match           13.6%;  Score 260;  DB 2;  Length 2531;
  Best Local Similarity  33.6%;  Pred. No. 1.92e-35;
  Matches   41;  Conservative   27;  Mismatches  43;  Indels  11;  Gaps  10;

Db    714 LSEVNECNSNPCIHGACR-DGLNGYKCDCAPGWSGTNCDINNNE-CESNPCVNGGTCKDM 771
          |:  : |: : | || |:  : :|: |:|  || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TS-GYVCTCREGFSGPNC--QTNI-NECASNPCLNQGTC-IDDVA-GYKCNCPLPYTGAT 825
           : :| | | || :|  |  : :: | |    | ::| | :  |:  | |:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CE 827
          |:
Qy    161 CD 162


RESULT   11
ENTRY            A31640     #type fragment
TITLE            epidermal growth factor-like protein slit - fruit fly
                   (Drosophila melanogaster) (fragment)
ORGANISM         #formal_name Drosophila melanogaster
DATE             28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
                   14-Aug-1998
ACCESSIONS       A31640
REFERENCE        A31640
   #authors      Rothberg, J.M.; Hartley, D.A.; Walther, Z.;
                   Artavanis-Tsakonas, S.
   #journal      Cell (1988) 55:1047-1059
   #title        slit: An EGF-homologous locus of D. melanogaster involved in
                   the development of the embryonic central nervous system.
   #cross-references MUID:89077533
   #accession    A31640
      ##molecule_type DNA
      ##residues      1-530 ##label ROT
      ##cross-references GB:M23543; NID:g340939; PID:g514357
GENETICS
   #gene         FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
   #introns      470/3
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         growth factor
FEATURE
   148-181           #domain EGF homology #label EGF
SUMMARY          #length 530  #checksum 6330

  Query Match           13.3%;  Score 254;  DB 2;  Length 530;
  Best Local Similarity  32.6%;  Pred. No. 4.27e-34;
  Matches   45;  Conservative   30;  Mismatches  51;  Indels  12;  Gaps  11;

Db     20 VRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEFMI-DACYGNPCRNN 78
          ::  ||: |::|    |  :: | : :|  : | |: |  |  |:    |:| || |  :
Qy     39 MQTGILPGCEPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VH 96

Db     79 ATCTVLEEGRFSCQCAPGYTGARCETNID--D-CLGEIKCQNNATC-IDGV-ESYKCECQ 133
            :|| ::   :|| |  |  |  |: : |  : |   ||| ::: | : || ::| |||
Qy     97 GTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQ-MIKC-KHGKCRLSGVGQPY-CECN 153

Db    134 PGFSGEFCDTKIQFCSPE 151
             :||:|: ||  |  |  |
Qy    154 SGFTGDSCDREIS-CRGE 170


RESULT   12
ENTRY            A36665     #type complete
TITLE            slit protein 1 precursor - fruit fly (Drosophila
                   melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change
                   24-Sep-1998
ACCESSIONS       A36665; S13523
REFERENCE        A36665
   #authors      Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.;
                   Artavanis-Tsakonas, S.
   #journal      Genes Dev. (1990) 4:2169-2187
   #title        slit: an extracellular protein necessary for development of
                   midline glia and commissural axon pathways contains both
                   EGF and LRR domains.
   #cross-references MUID:91099665
   #accession    A36665
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-1480 ##label ROT
      ##cross-references GB:X53959; NID:g8614; PID:g8615
GENETICS
   #gene         FlyBase:sli
      ##cross-references FlyBase:FBgn0003425
CLASSIFICATION   #superfamily proteoglycan amino-terminal homology; EGF
                   homology; leucine-rich alpha-2-glycoprotein repeat
                   homology; proteoglycan carboxyl-terminal homology
KEYWORDS         alternative splicing
FEATURE
   66-91             #domain proteoglycan amino-terminal homology #label
                       PAH1\
   101-124           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   125-148           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   149-172           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   173-196           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   197-220           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   228-272           #domain proteoglycan carboxyl-terminal homology #label
                       PCS1\
   288-313           #domain proteoglycan amino-terminal homology #label
                       PAH2\
   323-346           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6\
   347-370           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR7\
   371-394           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR8\
   395-418           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR9\
   419-442           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR10\
   450-494           #domain proteoglycan carboxyl-terminal homology #label
                       PCS2\
   512-537           #domain proteoglycan amino-terminal homology #label
                       PAH3\
   547-571           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR11\
   572-595           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR12\
   596-619           #domain leucine-rich alpha-2-glycoprotein repeat
```

```
                  homology #label LR13\
  620-643         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR14\
  651-695         #domain proteoglycan carboxyl-terminal homology #label
                  PCS3\
  708-733         #domain proteoglycan amino-terminal homology #label
                  PAH4\
  743-766         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR15\
  767-790         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR16\
  791-814         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR17\
  815-838         #domain leucine-rich alpha-2-glycoprotein repeat
                  homology #label LR18\
  846-890         #domain proteoglycan carboxyl-terminal homology #label
                  PCS4\
  1028-1061       #domain EGF homology #label EGF
SUMMARY         #length 1480  #molecular-weight 165751  #checksum 900

 Query Match           13.3%;  Score 254;  DB 2;  Length 1480;
 Best Local Similarity  32.6%;  Pred. No. 4.27e-34;
 Matches   45;  Conservative   30;  Mismatches  51;  Indels  12;  Gaps  11;

Db    900 VRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEFMI-DACYGNPCRNN 958
          ::  ||: |::|    |  :: | : :|  : | |: |  |  |:    |:| || |  :
Qy     39 MQTGILPGCEPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VH 96

Db    959 ATCTVLEEGRFSCQCAPGYTGARCETNID--D-CLGEIKCQNNATC-IDGV-ESYKCECQ 1013
          :||  ::   :|| |  |  |  |: : |  : |   ||| ::: | : || ::| |||
Qy     97 GTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQ-MIKC-KHGKCRLSGVGQPY-CECN 153

Db   1014 PGFSGEFCDTKIQFCSPE 1031
          :||:|: ||  |  |  |
Qy    154 SGFTGDSCDREIS-CRGE 170


RESULT   13
ENTRY           S42612     #type complete
TITLE           transmembrane protein precursor - zebra fish
ORGANISM        #formal_name Brachydanio rerio #common_name zebra fish
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  10-Jul-1998
ACCESSIONS      S42612
REFERENCE       S42612
   #authors     Bierkamp, C.; Campos-Ortega, J.A.
   #journal     Mech. Dev. (1993) 43:87-100
   #title       A zebrafish homologue of the Drosophila neurogenic gene Notch
                  and its pattern of transcription during early
                  embryogenesis.
   #cross-references MUID:94128602
   #accession   S42612
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2437 ##label BIE
      ##cross-references EMBL:X69088; NID:g433866; PID:g433867
CLASSIFICATION  #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology
FEATURE
   1915-1947         #domain ankyrin repeat homology #label AN1\
   1948-1980         #domain ankyrin repeat homology #label AN2\
   1982-2014         #domain ankyrin repeat homology #label AN3\
   2015-2047         #domain ankyrin repeat homology #label AN4\
   2048-2080         #domain ankyrin repeat homology #label AN5
SUMMARY         #length 2437  #molecular-weight 262306  #checksum 4021

  Query Match           13.2%;  Score 251;  DB 2;  Length 2437;
  Best Local Similarity  34.4%;  Pred. No. 2.00e-33;
  Matches   42;  Conservative   23;  Mismatches  45;  Indels  12;  Gaps  11;

Db    718 CSSNPCIHGSCLDQINS-YRCVCEAGWMGRNCDININE-CLSNPCVNGGTCKDMTS-GYL 774
          |  : | || |     | : | || ||||  ||   |: ||:| ||:| ||  : : :|
Qy     50 CHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYS 108

Db    775 CTCRAGFSGPNC--QMNI-NECASNPCLNQGSC-IDDVA-GFKCNCMLPYTGEVCENVLA 829
          | |  | :|  |  : :: | |    | ::| | :  |: : |:|   :||: |:  ::
Qy    109 CKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDSCDREIS 166

Db    830 PC 831
          |
Qy    167 -C 167


RESULT   14
ENTRY           A35672     #type complete
TITLE           crumbs protein - fruit fly (Drosophila melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A35672
REFERENCE       A35672
   #authors     Tepass, U.; Theres, C.; Knust, E.
   #journal     Cell (1990) 61:787-799
   #title       crumbs encodes an EGF-like protein expressed on apical
                  membranes of Drosophila epithelial cells and required for
                  organization of epithelia.
   #cross-references MUID:90263104
   #accession   A35672
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-2139 ##label TEP
      ##cross-references GB:M33753
      ##note        the authors translated the codon GGC for residue 1928 as
                     Cys, and TAT for residue 2023 as Gln
GENETICS
   #gene        FlyBase:crb
      ##cross-references FlyBase:FBgn0000368
CLASSIFICATION  #superfamily EGF homology
KEYWORDS        transmembrane protein
FEATURE
   691-722           #domain EGF homology #label EGF
SUMMARY         #length 2139  #molecular-weight 233619  #checksum 7230

  Query Match           12.9%;  Score 245;  DB 2;  Length 2139;
  Best Local Similarity  31.7%;  Pred. No. 4.36e-32;
  Matches   40;  Conservative   23;  Mismatches  57;  Indels   6;  Gaps   6;

Db    271 CLNDPCMGHGTC-SSSPEGYECRCTARYSGKNCQKDNGSPCAKNPCENGGSCLENSEGNY 329
          | :  | :|| | :||  |: | |     |  |:     ||  | | :| :||  :  :|
Qy     50 CHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSY 107

Db    330 QCFCDPNHSGQHCETEVNIHPLCQTNPCLNNGACVVIGGSGALTCECPKGYAGARCEVDT 389
          | |   |:|  |: | ::   ||   | ::| |  ::| |   |||  |::|  |: :
Qy    108 SCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCR-LSGVGQPYCECNSGFTGDSCDREI 165

Db    390 DECASQ 395
          | ::
Qy    166 S-CRGE 170


RESULT   15
ENTRY           A40043     #type complete
TITLE           notch protein homolog TAN-1 precursor - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
                  14-Aug-1998
ACCESSIONS      A40043
REFERENCE       A40043
   #authors     Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds,
                  T.C.; Smith, S.D.; Sklar, J.
   #journal     Cell (1991) 66:649-661
   #title       TAN-1, the human homolog of the Drosophila Notch gene, is
                  broken by chromosomal translocations in T lymphoblastic
                  neoplasms.
   #cross-references MUID:91347367
   #accession   A40043
```

```
     ##status       preliminary; nucleic acid sequence not shown; not
                      compared with conceptual translation
     ##molecule_type mRNA
     ##residues     1-2555 ##label ELL
     ##cross-references GB:M73980
CLASSIFICATION   #superfamily unassigned ankyrin repeat proteins; ankyrin
                  repeat homology; EGF homology
FEATURE
   1149-1180         #domain EGF homology #label EGF\
   1927-1959         #domain ankyrin repeat homology #label AN1\
   1960-1992         #domain ankyrin repeat homology #label AN2\
   1994-2026         #domain ankyrin repeat homology #label AN3\
   2027-2059         #domain ankyrin repeat homology #label AN4\
   2060-2092         #domain ankyrin repeat homology #label AN5
SUMMARY          #length 2555  #molecular-weight 272337  #checksum 463

  Query Match           12.9%;  Score 246;  DB 2;  Length 2555;
  Best Local Similarity  33.6%;  Pred. No. 2.61e-32;
  Matches   41;  Conservative   26;  Mismatches  44;  Indels  11;  Gaps  10;

Db    713 LSEVNECNSNPCVHGACR-DSLNGYKCDCDPGWSGTNCDINNNE-CESNPCVNGGTCKDM 770
          |:  : |: : | || |:  | :|: |:|: || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    771 TS-GIVCTCREGFSGPNC--QTNI-NECASNPCLNKGTC-IDDVA-GYKCNCLLPYTGAT 824
          : :  | | || :|  |  : :: | |    | : | | :  |:  | ||:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    825 CE 826
          |:
Qy    161 CD 162


Search completed: Fri May 28 09:41:04 1999
Job time : 40 secs.
```

MASPAR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Srch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:41:22 1999; MasPar time 8.63 Seconds
795.986 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-14
Description: (1-243) from US09191647.pep
Perfect Score: 1905
Sequence: 1 ILDVASLRQAPGENGTSFHG.....SSFVDEVEKVVKCGCARCAS 243

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 41.939; Variance 61.445; scale 0.683

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	292	15.3	1429	1	LI12_CAEEL	LIN-12 PROTEIN PRECURS	9.66e-48
2	274	14.4	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	3.65e-43
3	262	13.8	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	3.85e-40
4	260	13.6	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	1.22e-39
5	260	13.6	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	1.22e-39
6	254	13.3	1480	1	SLIT_DROME	SLIT PROTEIN PRECURSOR	3.86e-38
7	251	13.2	2437	1	NOTC_BRARE	NEUROGENIC LOCUS NOTCH	2.16e-37
8	245	12.9	2139	1	CRB_DROME	CRUMBS PROTEIN PRECURS	6.68e-36
9	246	12.9	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	3.78e-36
10	244	12.8	2703	1	NOTC_DROME	NEUROGENIC LOCUS NOTCH	1.18e-35
11	242	12.7	570	1	FBP3_STRPU	FIBROPELLIN C PRECURSO	3.69e-35
12	241	12.7	1964	1	NTC4_MOUSE	NEUROGENIC LOCUS NOTCH	6.52e-35
13	239	12.5	1295	1	GLP1_CAEEL	GLP-1 PROTEIN PRECURSO	2.03e-34
14	234	12.3	880	1	DL_DROME	NEUROGENIC LOCUS DELTA	3.45e-33
15	230	12.1	1064	1	FBP1_STRPU	FIBROPELLIN I PRECURSO	3.30e-32
16	231	12.1	4393	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	1.88e-32
17	215	11.3	714	1	DLL1_RAT	DELTA-LIKE PROTEIN 1 P	1.45e-28
18	213	11.2	722	1	DLL1_MOUSE	DELTA-LIKE PROTEIN 1 P	4.41e-28
19	211	11.1	723	1	DLL1_HUMAN	DELTA-LIKE PROTEIN 1 P	1.33e-27
20	212	11.1	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	7.67e-28
21	208	10.9	3707	1	PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	6.98e-27
22	203	10.7	1408	1	SERR_DROME	SERRATE PROTEIN PRECUR	1.09e-25
23	197	10.3	383	1	DLK_HUMAN	DELTA-LIKE PROTEIN PRE	2.87e-24
24	197	10.3	1268	1	PGCN_MOUSE	NEUROCAN CORE PROTEIN	2.87e-24
25	195	10.2	1257	1	PGCN_RAT	NEUROCAN CORE PROTEIN	8.51e-24
26	188	9.9	3396	1	PGCV_HUMAN	VERSICAN CORE PROTEIN	3.72e-22
27	185	9.7	1959	1	AGRI_RAT	AGRIN PRECURSOR.	1.86e-21
28	185	9.7	3358	1	PGCV_MOUSE	VERSICAN CORE PROTEIN	1.86e-21
29	183	9.6	862	1	PGCV_MACNE	VERSICAN CORE PROTEIN	5.40e-21
30	179	9.4	385	1	DLK_MOUSE	DELTA-LIKE PROTEIN PRE	4.52e-20
31	177	9.3	1394	1	TGFB_HUMAN	LATENT TRANSFORMING GR	1.30e-19
32	177	9.3	1712	1	TGFB_RAT	LATENT TRANSFORMING GR	1.30e-19
33	178	9.3	2871	1	FBN1_BOVIN	FIBRILLIN 1 PRECURSOR	7.68e-20
34	177	9.3	2871	1	FBN1_HUMAN	FIBRILLIN 1 PRECURSOR.	1.30e-19
35	176	9.2	1955	1	AGRI_CHICK	AGRIN PRECURSOR.	2.21e-19
36	175	9.2	3562	1	PGCV_CHICK	VERSICAN CORE PROTEIN	3.74e-19
37	172	9.0	2911	1	FBN2_HUMAN	FIBRILLIN 2 PRECURSOR.	1.81e-18
38	170	8.9	2871	1	FBN1_MOUSE	FIBRILLIN 1 PRECURSOR.	5.14e-18
39	166	8.7	379	1	CYR6_MOUSE	CYR61 PROTEIN PRECURSO	4.11e-17
40	165	8.7	458	1	PRTC_RABIT	VITAMIN-K DEPENDENT PR	6.90e-17
41	162	8.5	375	1	CE10_CHICK	CEF-10 PROTEIN PRECURS	3.24e-16
42	161	8.5	2907	1	FBN2_MOUSE	FIBRILLIN 2 PRECURSOR.	5.41e-16
43	156	8.2	1328	1	AGRI_DISOM	AGRIN (FRAGMENT).	6.95e-15
44	153	8.0	381	1	CYR6_HUMAN	CYR61 PROTEIN PRECURSO	3.17e-14
45	151	7.9	515	1	APX1_CAEEL	APX-1 PROTEIN PRECURSO	8.67e-14

ALIGNMENTS

RESULT 1
ID LI12_CAEEL STANDARD; PRT; 1429 AA.
AC P14585;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE LIN-12 PROTEIN PRECURSOR.
GN LIN-12 OR R107.8.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 88334747.
RA YOCHEM J., WESTON K., GREENWALD I.;
RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
RT protein with overall similarity to Drosophila Notch.";
RL NATURE 335:547-550(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
CC -!- FUNCTION: LIN-12 IS IS INVOLVED IN SEVERAL CELL FATES DECISIONS
CC THAT REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
CC -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC ---

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12069; G156358; -.
DR   EMBL; Z14092; E1348691; -.
DR   PIR; S06434; S06434.
DR   WORMPEP; R107.8; CE00274.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 12.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 13.
DR   PFAM; PF00023; ank; 4.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1     15       POTENTIAL.
FT   CHAIN       16   1429       LIN-12 PROTEIN.
FT   DOMAIN      16    908       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   909    931       POTENTIAL.
FT   DOMAIN     932   1429       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24    618       13 X EGF-TYPE REPEATS.
FT   DOMAIN     631    750       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1046   1266       6 X ANK MOTIF REPEATS.
FT   DOMAIN      20     61       EGF-LIKE 1.
FT   DOMAIN     114    150       EGF-LIKE 2.
FT   DOMAIN     152    190       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     201    246       EGF-LIKE 4.
FT   DOMAIN     250    285       EGF-LIKE 5.
FT   DOMAIN     287    323       EGF-LIKE 6.
FT   DOMAIN     323    363       EGF-LIKE 7.
FT   DOMAIN     365    402       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    441       EGF-LIKE 9.
FT   DOMAIN     449    492       EGF-LIKE 10.
FT   DOMAIN     503    541       EGF-LIKE 11.
FT   DOMAIN     543    579       EGF-LIKE 12.
FT   DOMAIN     582    619       EGF-LIKE 13.
FT   REPEAT     635    669       LIN/NOTCH 1.
FT   REPEAT     670    710       LIN/NOTCH 2.
FT   REPEAT     711    750       LIN/NOTCH 3.
FT   REPEAT    1046   1078       ANK MOTIF 1.
FT   REPEAT    1079   1119       ANK MOTIF 2.
FT   REPEAT    1120   1152       ANK MOTIF 3.
FT   REPEAT    1153   1188       ANK MOTIF 4.
FT   REPEAT    1189   1232       ANK MOTIF 5.
FT   REPEAT    1233   1266       ANK MOTIF 6.
FT   DISULFID    24     35       BY SIMILARITY.
FT   DISULFID    29     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID   118    129       BY SIMILARITY.
FT   DISULFID   123    138       BY SIMILARITY.
FT   DISULFID   140    149       BY SIMILARITY.
FT   DISULFID   156    169       BY SIMILARITY.
FT   DISULFID   163    178       BY SIMILARITY.
FT   DISULFID   180    189       BY SIMILARITY.
FT   DISULFID   205    227       BY SIMILARITY.
FT   DISULFID   221    234       BY SIMILARITY.
FT   DISULFID   236    245       BY SIMILARITY.
FT   DISULFID   254    264       BY SIMILARITY.
FT   DISULFID   259    273       BY SIMILARITY.
FT   DISULFID   275    284       BY SIMILARITY.
FT   DISULFID   291    302       BY SIMILARITY.
FT   DISULFID   296    311       BY SIMILARITY.
FT   DISULFID   313    322       BY SIMILARITY.
FT   DISULFID   327    339       BY SIMILARITY.
FT   DISULFID   334    351       BY SIMILARITY.
FT   DISULFID   353    362       BY SIMILARITY.
FT   DISULFID   369    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    429       BY SIMILARITY.
FT   DISULFID   431    440       BY SIMILARITY.
FT   DISULFID   507    518       BY SIMILARITY.
FT   DISULFID   512    529       BY SIMILARITY.
FT   DISULFID   531    540       BY SIMILARITY.
FT   DISULFID   547    558       BY SIMILARITY.
FT   DISULFID   552    567       BY SIMILARITY.
FT   DISULFID   569    578       BY SIMILARITY.
FT   DISULFID   586    597       BY SIMILARITY.
FT   DISULFID   591    607       BY SIMILARITY.
FT   DISULFID   609    618       BY SIMILARITY.
FT   CARBOHYD    41     41       POTENTIAL.
FT   CARBOHYD   165    165       POTENTIAL.
FT   CARBOHYD   194    194       POTENTIAL.
FT   CARBOHYD   378    378       POTENTIAL.
FT   CARBOHYD   515    515       POTENTIAL.
FT   CARBOHYD   623    623       POTENTIAL.
FT   CARBOHYD   751    751       POTENTIAL.
FT   CARBOHYD   754    754       POTENTIAL.
FT   CARBOHYD   900    900       POTENTIAL.
SQ   SEQUENCE   1429 AA;  157115 MW;  CFD2CCA4 CRC32;

  Query Match          15.3%;  Score 292;  DB 1;  Length 1429;
  Best Local Similarity  40.5%;  Pred. No. 9.66e-48;
  Matches   47;  Conservative   24;  Mismatches  35;  Indels  10;  Gaps   9;

Db    333 ICNHGTCIDSPLSEKAFECQCEPGYEGILCEQDKNE-CLSENMCLNNGTCVNLPG-SFRC 390
          :| || | : | |: :| |:|| | | ||:|  |: ||: |:|:: |||: : : |: |
Qy     54 VCAHGCC-Q-PSSQSGFTCECEEGWMGPLCDQRTNDPCLG-NKCVH-GTCLPINAFSYSC 109

Db    391 DCARGFGGKWCDEP--L-NMCQDFHCENDGTCMHTSDHSPVCQCKNGFIGKRCEKE 443
           |  | ||  |||   | | || : | : | |  ::   | |:|::|| |  |::|
Qy    110 KCLEGHGGVLCDEEEDLFNPCQMIKCKH-GKCRLSGVGQPYCECNSGFTGDSCDRE 164


RESULT    2
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   LARDELLI M., DALSTRAND J., LENDAHL U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   MECH. DEV. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X74760; G483581; -.
DR   MGD; MGI:99460; NOTCH3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 33.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN.
FT   DOMAIN       1   1643       EXTRACELLULAR.
FT   TRANSMEM  1644   1664       POTENTIAL.
FT   DOMAIN    1665   2318       CYTOPLASMIC.
FT   DOMAIN      39   1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN    1388   1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN    1784   1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN    2242   2261       PEST.
FT   DOMAIN      39     78       EGF-LIKE 1.
FT   DOMAIN      79    119       EGF-LIKE 2.
FT   DOMAIN     120    157       EGF-LIKE 3.
FT   DOMAIN     159    196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235       EGF-LIKE 5.
FT   DOMAIN     237    273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313       EGF-LIKE 7.
FT   DOMAIN     315    351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390       EGF-LIKE 9.
FT   DOMAIN     392    430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731       EGF-LIKE 18.
FT   DOMAIN     735    771       EGF-LIKE 19.
FT   DOMAIN     772    809       EGF-LIKE 20.
FT   DOMAIN     811    848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961       EGF-LIKE 24.
FT   DOMAIN     963    999       EGF-LIKE 25.
FT   DOMAIN    1001   1035       EGF-LIKE 26.
FT   DOMAIN    1037   1083       EGF-LIKE 27.
FT   DOMAIN    1085   1121       EGF-LIKE 28.
FT   DOMAIN    1123   1159       EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245       EGF-LIKE 31.
FT   DOMAIN    1247   1288       EGF-LIKE 32.
FT   DOMAIN    1290   1326       EGF-LIKE 33.
FT   DOMAIN    1336   1374       EGF-LIKE 34.
FT   REPEAT    1388   1428       LIN/NOTCH 1.
FT   REPEAT    1429   1467       LIN/NOTCH 2.
FT   REPEAT    1468   1503       LIN/NOTCH 3.
FT   REPEAT    1784   1816       CDC10/SWI6 1.
FT   REPEAT    1817   1865       CDC10/SWI6 2.
FT   REPEAT    1866   1898       CDC10/SWI6 3.
FT   REPEAT    1899   1932       CDC10/SWI6 4.
FT   REPEAT    1933   1965       CDC10/SWI6 5.
FT   REPEAT    1966   1998       CDC10/SWI6 6.
FT   DISULFID    43     55       BY SIMILARITY.
FT   DISULFID    49     66       BY SIMILARITY.
FT   DISULFID    68     77       BY SIMILARITY.
FT   DISULFID    83     94       BY SIMILARITY.
FT   DISULFID    88    107       BY SIMILARITY.
FT   DISULFID   109    118       BY SIMILARITY.
FT   DISULFID   124    135       BY SIMILARITY.
FT   DISULFID   129    145       BY SIMILARITY.
FT   DISULFID   147    156       BY SIMILARITY.
FT   DISULFID   163    175       BY SIMILARITY.
FT   DISULFID   169    184       BY SIMILARITY.
FT   DISULFID   186    195       BY SIMILARITY.
FT   DISULFID   202    213       BY SIMILARITY.
FT   DISULFID   207    223       BY SIMILARITY.
FT   DISULFID   225    234       BY SIMILARITY.
FT   DISULFID   241    252       BY SIMILARITY.
FT   DISULFID   246    261       BY SIMILARITY.
FT   DISULFID   263    272       BY SIMILARITY.
FT   DISULFID   279    292       BY SIMILARITY.
FT   DISULFID   286    301       BY SIMILARITY.
FT   DISULFID   303    312       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   324    339       BY SIMILARITY.
FT   DISULFID   341    350       BY SIMILARITY.
FT   DISULFID   356    367       BY SIMILARITY.
FT   DISULFID   361    378       BY SIMILARITY.
FT   DISULFID   380    389       BY SIMILARITY.
FT   DISULFID   396    409       BY SIMILARITY.
FT   DISULFID   403    418       BY SIMILARITY.
FT   DISULFID   420    429       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   441    456       BY SIMILARITY.
FT   DISULFID   458    467       BY SIMILARITY.
FT   DISULFID   474    485       BY SIMILARITY.
FT   DISULFID   479    494       BY SIMILARITY.
FT   DISULFID   496    505       BY SIMILARITY.
FT   DISULFID   512    523       BY SIMILARITY.
FT   DISULFID   517    532       BY SIMILARITY.
FT   DISULFID   534    543       BY SIMILARITY.
FT   DISULFID   550    560       BY SIMILARITY.
FT   DISULFID   555    569       BY SIMILARITY.
FT   DISULFID   571    580       BY SIMILARITY.
FT   DISULFID   587    598       BY SIMILARITY.
FT   DISULFID   592    607       BY SIMILARITY.
FT   DISULFID   609    618       BY SIMILARITY.
FT   DISULFID   625    635       BY SIMILARITY.
FT   DISULFID   630    644       BY SIMILARITY.
FT   DISULFID   646    655       BY SIMILARITY.
FT   DISULFID   662    673       BY SIMILARITY.
FT   DISULFID   667    682       BY SIMILARITY.
FT   DISULFID   684    693       BY SIMILARITY.
FT   DISULFID   700    710       BY SIMILARITY.
FT   DISULFID   705    719       BY SIMILARITY.
FT   DISULFID   721    730       BY SIMILARITY.
FT   DISULFID   739    750       BY SIMILARITY.
FT   DISULFID   744    759       BY SIMILARITY.
FT   DISULFID   761    770       BY SIMILARITY.
FT   DISULFID   776    787       BY SIMILARITY.
FT   DISULFID   781    797       BY SIMILARITY.
FT   DISULFID   799    808       BY SIMILARITY.
FT   DISULFID   815    827       BY SIMILARITY.
FT   DISULFID   821    836       BY SIMILARITY.
FT   DISULFID   838    847       BY SIMILARITY.
FT   DISULFID   854    865       BY SIMILARITY.
FT   DISULFID   859    874       BY SIMILARITY.
FT   DISULFID   876    885       BY SIMILARITY.
FT   DISULFID   892    902       BY SIMILARITY.
FT   DISULFID   897    911       BY SIMILARITY.
FT   DISULFID   913    922       BY SIMILARITY.
FT   DISULFID   929    940       BY SIMILARITY.
FT   DISULFID   934    949       BY SIMILARITY.
FT   DISULFID   951    960       BY SIMILARITY.
FT   DISULFID   967    978       BY SIMILARITY.
FT   DISULFID   972    987       BY SIMILARITY.
FT   DISULFID   989    998       BY SIMILARITY.
FT   DISULFID  1005   1016       BY SIMILARITY.
FT   DISULFID  1010   1023       BY SIMILARITY.
FT   DISULFID  1025   1034       BY SIMILARITY.
FT   DISULFID  1041   1062       BY SIMILARITY.
FT   DISULFID  1056   1071       BY SIMILARITY.
FT   DISULFID  1073   1082       BY SIMILARITY.
```

```
FT   DISULFID   1089   1100       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
FT   DISULFID   1111   1120       BY SIMILARITY.
FT   DISULFID   1127   1138       BY SIMILARITY.
FT   DISULFID   1132   1147       BY SIMILARITY.
FT   DISULFID   1149   1158       BY SIMILARITY.
FT   DISULFID   1165   1183       BY SIMILARITY.
FT   DISULFID   1177   1192       BY SIMILARITY.
FT   DISULFID   1194   1203       BY SIMILARITY.
FT   DISULFID   1210   1223       BY SIMILARITY.
FT   DISULFID   1215   1233       BY SIMILARITY.
FT   DISULFID   1235   1244       BY SIMILARITY.
FT   DISULFID   1251   1262       BY SIMILARITY.
FT   DISULFID   1256   1276       BY SIMILARITY.
FT   DISULFID   1278   1287       BY SIMILARITY.
FT   DISULFID   1294   1305       BY SIMILARITY.
FT   DISULFID   1299   1314       BY SIMILARITY.
FT   DISULFID   1316   1325       BY SIMILARITY.
FT   DISULFID   1340   1351       BY SIMILARITY.
FT   DISULFID   1345   1362       BY SIMILARITY.
FT   DISULFID   1364   1373       BY SIMILARITY.

[illegible]e: remainder of annotations omitted.

  Query Match          14.4%;  Score 274;  DB 1;  Length 2318;
  Best Local Similarity  36.2%;  Pred. No. 3.65e-43;
  Matches   47;  Conservative   21;  Mismatches  50;  Indels  12;  Gaps  11;

Db    122 DPCVSRPCVHGAPCSVGPDGRFACACPPGYQGQSC-QSDIDECRSGTTCRHGGTCLNTPG 180
          :||  : | ||  |  :::: |:| |  |  |  | |   | |  |  | || |||   :
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCL-GNKCVHG-TCLPINA 104

Db    181 -SFRCQCPLGYTGLLC--ENPVV-PCAPSPCRNGGTCRQSS-DVTYDCACLPGFEGQNCE 235
           |: | |  |  |:||  |: :  ||    |::|  || |:    | | | :|| |::|:
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCD 162

Db    236 VNVDDCPGHR 245
           ::  | | |
Qy    163 REIS-CRGER 171


RESULT    3
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
[illegible]  XENOPUS LAEVIS (AFRICAN CLAWED FROG).
[illegible]  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
     MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RT   "Xotch, the Xenopus homolog of Drosophila notch.";
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN      20   1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1729   1750       POTENTIAL.
FT   DOMAIN    1751   2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     57       EGF-LIKE 1.
FT   DOMAIN      58     99       EGF-LIKE 2.
FT   DOMAIN     102    140       EGF-LIKE 3.
FT   DOMAIN     141    177       EGF-LIKE 4.
FT   DOMAIN     179    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     217    254       EGF-LIKE 6.
FT   DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     371    409       EGF-LIKE 10.
FT   DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     489    525       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     527    563       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     565    600       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    638       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     640    675       EGF-LIKE 17.
FT   DOMAIN     677    713       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    750       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     752    788       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     790    826       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     828    866       EGF-LIKE 22.
FT   DOMAIN     868    904       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     906    942       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     944    980       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     982   1018       EGF-LIKE 26.
FT   DOMAIN    1020   1056       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1058   1094       EGF-LIKE 28.
FT   DOMAIN    1096   1142       EGF-LIKE 29.
FT   DOMAIN    1144   1180       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1182   1218       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1220   1264       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1266   1304       EGF-LIKE 33.
FT   DOMAIN    1306   1346       EGF-LIKE 34.
FT   DOMAIN    1347   1383       EGF-LIKE 35.
FT   DOMAIN    1386   1424       EGF-LIKE 36.
FT   DOMAIN    1441   1560       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1441   1478       LIN/NOTCH 1.
FT   REPEAT    1479   1520       LIN/NOTCH 2.
FT   REPEAT    1521   1560       LIN/NOTCH 3.
FT   DOMAIN    1871   2083       6 X ANK MOTIF REPEATS.
FT   DISULFID    22     35       BY SIMILARITY.
FT   DISULFID    29     45       BY SIMILARITY.
FT   DISULFID    47     56       BY SIMILARITY.
FT   DISULFID    62     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    128       BY SIMILARITY.
FT   DISULFID   130    139       BY SIMILARITY.
FT   DISULFID   145    156       BY SIMILARITY.
FT   DISULFID   150    165       BY SIMILARITY.
```

```
FT   DISULFID   167   176       BY SIMILARITY.
FT   DISULFID   183   194       BY SIMILARITY.
FT   DISULFID   188   203       BY SIMILARITY.
FT   DISULFID   205   214       BY SIMILARITY.
FT   DISULFID   221   232       BY SIMILARITY.
FT   DISULFID   226   242       BY SIMILARITY.
FT   DISULFID   244   253       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   265   280       BY SIMILARITY.
FT   DISULFID   282   291       BY SIMILARITY.
FT   DISULFID   298   311       BY SIMILARITY.
FT   DISULFID   305   320       BY SIMILARITY.
FT   DISULFID   322   331       BY SIMILARITY.
FT   DISULFID   338   349       BY SIMILARITY.
FT   DISULFID   343   358       BY SIMILARITY.
FT   DISULFID   360   369       BY SIMILARITY.
FT   DISULFID   375   386       BY SIMILARITY.
FT   DISULFID   380   397       BY SIMILARITY.
FT   DISULFID   399   408       BY SIMILARITY.
FT   DISULFID   415   428       BY SIMILARITY.
FT   DISULFID   422   437       BY SIMILARITY.
FT   DISULFID   439   448       BY SIMILARITY.
FT   DISULFID   455   466       BY SIMILARITY.
FT   DISULFID   460   475       BY SIMILARITY.
FT   DISULFID   477   486       BY SIMILARITY.
FT   DISULFID   493   504       BY SIMILARITY.
FT   DISULFID   498   513       BY SIMILARITY.
FT   DISULFID   515   524       BY SIMILARITY.
FT   DISULFID   531   542       BY SIMILARITY.
FT   DISULFID   536   551       BY SIMILARITY.
FT   DISULFID   553   562       BY SIMILARITY.
FT   DISULFID   569   579       BY SIMILARITY.
FT   DISULFID   574   588       BY SIMILARITY.
FT   DISULFID   590   599       BY SIMILARITY.
FT   DISULFID   606   617       BY SIMILARITY.
FT   DISULFID   611   626       BY SIMILARITY.
FT   DISULFID   628   637       BY SIMILARITY.
FT   DISULFID   644   654       BY SIMILARITY.
FT   DISULFID   649   663       BY SIMILARITY.
FT   DISULFID   665   674       BY SIMILARITY.
FT   DISULFID   681   692       BY SIMILARITY.
FT   DISULFID   686   701       BY SIMILARITY.
FT   DISULFID   703   712       BY SIMILARITY.
FT   DISULFID   719   729       BY SIMILARITY.
FT   DISULFID   724   738       BY SIMILARITY.
FT   DISULFID   740   749       BY SIMILARITY.
FT   DISULFID   756   767       BY SIMILARITY.
FT   DISULFID   761   776       BY SIMILARITY.
FT   DISULFID   778   787       BY SIMILARITY.
FT   DISULFID   794   805       BY SIMILARITY.
FT   DISULFID   799   814       BY SIMILARITY.
FT   DISULFID   816   825       BY SIMILARITY.
FT   DISULFID   832   843       BY SIMILARITY.
FT   DISULFID   837   854       BY SIMILARITY.
FT   DISULFID   856   865       BY SIMILARITY.
FT   DISULFID   872   883       BY SIMILARITY.
FT   DISULFID   877   892       BY SIMILARITY.
FT   DISULFID   894   903       BY SIMILARITY.
FT   DISULFID   910   921       BY SIMILARITY.
FT   DISULFID   915   930       BY SIMILARITY.
FT   DISULFID   932   941       BY SIMILARITY.
FT   DISULFID   986   997       BY SIMILARITY.
FT   DISULFID   991  1006       BY SIMILARITY.
FT   DISULFID  1008  1017       BY SIMILARITY.
FT   DISULFID  1024  1035       BY SIMILARITY.
FT   DISULFID  1029  1044       BY SIMILARITY.
FT   DISULFID  1046  1055       BY SIMILARITY.
FT   DISULFID  1062  1073       BY SIMILARITY.
FT   DISULFID  1067  1082       BY SIMILARITY.
FT   DISULFID  1084  1093       BY SIMILARITY.
FT   DISULFID  1100  1121       BY SIMILARITY.
FT   DISULFID  1115  1130       BY SIMILARITY.
FT   DISULFID  1132  1141       BY SIMILARITY.
FT   DISULFID  1148  1159       BY SIMILARITY.
FT   DISULFID  1153  1168       BY SIMILARITY.
FT   DISULFID  1170  1179       BY SIMILARITY.
FT   DISULFID  1186  1197       BY SIMILARITY.
FT   DISULFID  1191  1206       BY SIMILARITY.
FT   DISULFID  1208  1217       BY SIMILARITY.
FT   DISULFID  1224  1243       BY SIMILARITY.
FT   DISULFID  1237  1252       BY SIMILARITY.
FT   DISULFID  1254  1263       BY SIMILARITY.
FT   DISULFID  1270  1283       BY SIMILARITY.
FT   DISULFID  1275  1292       BY SIMILARITY.
FT   DISULFID  1294  1303       BY SIMILARITY.
FT   DISULFID  1310  1321       BY SIMILARITY.
FT   DISULFID  1315  1333       BY SIMILARITY.
FT   DISULFID  1335  1344       BY SIMILARITY.
FT   DISULFID  1351  1362       BY SIMILARITY.
FT   DISULFID  1356  1371       BY SIMILARITY.
FT   DISULFID  1373  1382       BY SIMILARITY.
FT   DISULFID  1390  1401       BY SIMILARITY.
FT   DISULFID  1395  1412       BY SIMILARITY.
FT   DISULFID  1414  1423       BY SIMILARITY.
FT   CARBOHYD   462   462       POTENTIAL.
FT   CARBOHYD   887   887       POTENTIAL.
...
Note: remainder of annotations omitted.

  Query Match          13.8%;  Score 262;  DB 1;  Length 2524;
  Best Local Similarity  31.9%;  Pred. No. 3.85e-40;
  Matches   43;  Conservative   31;  Mismatches  49;  Indels  12;  Gaps  11;

Db    713 LSEVNECNSNPCIHGACHDGV-NGYKCDCEAGWSGSNCDINNNE-CESNPCMNGGTCKDM 770
          |:  : |: : | || |: :  :|: |:|| || |: ||   |: | :| |::| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    771 TG-AYICTCKAGFSGPNC--QTNI-NECSSNPCLNHGTC-IDDVA-GYKCNCMLPYTGAI 824
           : :| | |  | :|  |  : :: | |    | :|| | :  |:  | |:|   :||
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    825 CEAVLAPCAGSPCKN 839
          |:  :: | |   ::
Qy    161 CDREIS-CRGERIRD 174


RESULT    4
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   DEVELOPMENT 115:737-744(1992).
```

```
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1726   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN    1449   1462       CYS-RICH.
FT   DOMAIN    1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1445   1480       LIN/NOTCH 1.
FT   REPEAT    1481   1522       LIN/NOTCH 2.
FT   REPEAT    1523   1562       LIN/NOTCH 3.
FT   DOMAIN    1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT    1865   1910       ANK MOTIF 1.
FT   REPEAT    1912   1942       ANK MOTIF 2.
FT   REPEAT    1944   1975       ANK MOTIF 3.
FT   REPEAT    1978   2009       ANK MOTIF 4.
FT   REPEAT    2011   2042       ANK MOTIF 5.
FT   REPEAT    2044   2075       ANK MOTIF 6.
FT   CARBOHYD   888    888       POTENTIAL.
FT   CARBOHYD   959    959       POTENTIAL.
FT   CARBOHYD  1179   1179       POTENTIAL.
FT   CARBOHYD  1241   1241       POTENTIAL.
FT   CARBOHYD  1489   1489       POTENTIAL.
FT   CARBOHYD  1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match          13.6%;  Score 260;  DB 1;  Length 2531;
  Best Local Similarity  33.6%;  Pred. No. 1.22e-39;
  Matches   41;  Conservative   27;  Mismatches  43;  Indels  11;  Gaps  10;

Db    714 LSEVNECNSNPCIHGACR-DGLNGYKCDCAPGWSGTNCDINNNE-CESNPCVNGGTCKDM 771
          |:  : |: : | || |:  : :|: |:|  || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TS-GYVCTCREGFSGPNC--QTNI-NECASNPCLNQGTC-IDDVA-GYKCNCPLPYTGAT 825
           : :| | | || :|  |  : :: | |    | ::| | :  |:  | |:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CE 827
          |:
Qy    161 CD 162


RESULT    5
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   WEINMASTER G., ROBERTS V.J., LEMKE G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   DEVELOPMENT 113:199-205(1991).
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57405; G57635; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN      19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1724   1746       POTENTIAL.
FT   DOMAIN    1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    867       EGF-LIKE 22.
```

```
FT   DOMAIN      869    905      EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      907    943      EGF-LIKE 24.
FT   DOMAIN      945    981      EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      983   1019      EGF-LIKE 26.
FT   DOMAIN     1021   1057      EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1059   1095      EGF-LIKE 28.
FT   DOMAIN     1097   1143      EGF-LIKE 29.
FT   DOMAIN     1145   1181      EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1183   1219      EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1265      EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1267   1305      EGF-LIKE 33.
FT   DOMAIN     1307   1346      EGF-LIKE 34.
FT   DOMAIN     1348   1384      EGF-LIKE 35.
FT   DOMAIN     1387   1426      EGF-LIKE 36.
FT   DOMAIN     1449   1462      CYS-RICH.
FT   DOMAIN     1865   2076      6 X ANK MOTIF REPEATS.
FT   REPEAT     1865   1910      ANK MOTIF 1.
FT   REPEAT     1912   1942      ANK MOTIF 2.
FT   REPEAT     1944   1975      ANK MOTIF 3.
FT   REPEAT     1978   2009      ANK MOTIF 4.
FT   REPEAT     2011   2042      ANK MOTIF 5.
FT   REPEAT     2044   2076      ANK MOTIF 6.
FT   DISULFID     24     37      BY SIMILARITY.
FT   DISULFID     31     46      BY SIMILARITY.
FT   DISULFID     48     57      BY SIMILARITY.
FT   DISULFID     63     74      BY SIMILARITY.
FT   DISULFID     68     87      BY SIMILARITY.
FT   DISULFID     89     98      BY SIMILARITY.
FT   DISULFID    106    117      BY SIMILARITY.
FT   DISULFID    111    127      BY SIMILARITY.
FT   DISULFID    129    138      BY SIMILARITY.
FT   DISULFID    144    155      BY SIMILARITY.
FT   DISULFID    149    164      BY SIMILARITY.
FT   DISULFID    166    175      BY SIMILARITY.
FT   DISULFID    182    195      BY SIMILARITY.
FT   DISULFID    189    204      BY SIMILARITY.
FT   DISULFID    206    215      BY SIMILARITY.
FT   DISULFID    222    233      BY SIMILARITY.
FT   DISULFID    227    243      BY SIMILARITY.
FT   DISULFID    245    254      BY SIMILARITY.
FT   DISULFID    261    272      BY SIMILARITY.
FT   DISULFID    266    281      BY SIMILARITY.
FT   DISULFID    283    292      BY SIMILARITY.
FT   DISULFID    299    312      BY SIMILARITY.
FT   DISULFID    306    321      BY SIMILARITY.
FT   DISULFID    323    332      BY SIMILARITY.
FT   DISULFID    339    350      BY SIMILARITY.
FT   DISULFID    344    359      BY SIMILARITY.
FT   DISULFID    361    370      BY SIMILARITY.
FT   DISULFID    376    387      BY SIMILARITY.
FT   DISULFID    381    398      BY SIMILARITY.
FT   DISULFID    400    409      BY SIMILARITY.
FT   DISULFID    416    429      BY SIMILARITY.
FT   DISULFID    423    438      BY SIMILARITY.
FT   DISULFID    440    449      BY SIMILARITY.
FT   DISULFID    456    467      BY SIMILARITY.
FT   DISULFID    461    476      BY SIMILARITY.
FT   DISULFID    478    487      BY SIMILARITY.
FT   DISULFID    494    505      BY SIMILARITY.
FT   DISULFID    499    514      BY SIMILARITY.
FT   DISULFID    516    525      BY SIMILARITY.
FT   DISULFID    532    543      BY SIMILARITY.
FT   DISULFID    537    552      BY SIMILARITY.
FT   DISULFID    554    563      BY SIMILARITY.
FT   DISULFID    570    580      BY SIMILARITY.
FT   DISULFID    575    589      BY SIMILARITY.
FT   DISULFID    591    600      BY SIMILARITY.
FT   DISULFID    607    618      BY SIMILARITY.
FT   DISULFID    612    627      BY SIMILARITY.
FT   DISULFID    629    638      BY SIMILARITY.
FT   DISULFID    645    655      BY SIMILARITY.
FT   DISULFID    650    664      BY SIMILARITY.
FT   DISULFID    666    675      BY SIMILARITY.
FT   DISULFID    682    693      BY SIMILARITY.
FT   DISULFID    687    702      BY SIMILARITY.
FT   DISULFID    704    713      BY SIMILARITY.
FT   DISULFID    720    730      BY SIMILARITY.
FT   DISULFID    725    739      BY SIMILARITY.
FT   DISULFID    741    750      BY SIMILARITY.
FT   DISULFID    757    768      BY SIMILARITY.
FT   DISULFID    762    777      BY SIMILARITY.
FT   DISULFID    779    788      BY SIMILARITY.
FT   DISULFID    795    806      BY SIMILARITY.
FT   DISULFID    800    815      BY SIMILARITY.
FT   DISULFID    817    826      BY SIMILARITY.
FT   DISULFID    833    844      BY SIMILARITY.
FT   DISULFID    838    855      BY SIMILARITY.
FT   DISULFID    857    866      BY SIMILARITY.
FT   DISULFID    873    884      BY SIMILARITY.
FT   DISULFID    878    893      BY SIMILARITY.
FT   DISULFID    895    904      BY SIMILARITY.
FT   DISULFID    911    922      BY SIMILARITY.
FT   DISULFID    916    931      BY SIMILARITY.
FT   DISULFID    933    942      BY SIMILARITY.
FT   DISULFID    987    998      BY SIMILARITY.
FT   DISULFID    992   1007      BY SIMILARITY.
FT   DISULFID   1009   1018      BY SIMILARITY.
FT   DISULFID   1025   1036      BY SIMILARITY.
FT   DISULFID   1030   1045      BY SIMILARITY.
FT   DISULFID   1047   1056      BY SIMILARITY.
FT   DISULFID   1063   1074      BY SIMILARITY.
FT   DISULFID   1068   1083      BY SIMILARITY.
FT   DISULFID   1085   1094      BY SIMILARITY.
FT   DISULFID   1101   1122      BY SIMILARITY.
FT   DISULFID   1116   1131      BY SIMILARITY.
FT   DISULFID   1133   1142      BY SIMILARITY.
FT   DISULFID   1149   1160      BY SIMILARITY.
FT   DISULFID   1154   1169      BY SIMILARITY.
FT   DISULFID   1171   1180      BY SIMILARITY.
FT   DISULFID   1187   1198      BY SIMILARITY.
FT   DISULFID   1192   1207      BY SIMILARITY.
FT   DISULFID   1209   1218      BY SIMILARITY.
FT   DISULFID   1225   1244      BY SIMILARITY.
FT   DISULFID   1238   1253      BY SIMILARITY.
FT   DISULFID   1255   1264      BY SIMILARITY.
FT   DISULFID   1271   1284      BY SIMILARITY.
FT   DISULFID   1276   1293      BY SIMILARITY.
FT   DISULFID   1295   1304      BY SIMILARITY.
FT   DISULFID   1311   1322      BY SIMILARITY.
FT   DISULFID   1316   1334      BY SIMILARITY.
FT   DISULFID   1336   1345      BY SIMILARITY.
FT   DISULFID   1352   1363      BY SIMILARITY.
FT   DISULFID   1357   1372      BY SIMILARITY.
FT   DISULFID   1374   1383      BY SIMILARITY.
FT   DISULFID   1391   1403      BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          13.6%;  Score 260;  DB 1;  Length 2531;
  Best Local Similarity  33.6%;  Pred. No. 1.22e-39;
  Matches   41;  Conservative   27;  Mismatches  43;  Indels  11;  Gaps  10;

Db    714 LSEVNECNSNPCIHGACR-DGLNGYKCDCAPGWSGTNCDINNNE-CESNPCVNGGTCKDM 771
          |: : |: : | || |:  : :|: |:|  || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TS-GYVCTCREGFSGPNC--QTNI-NECASNPCLNQGTC-IDDVA-GYKCNCPLPYTGAT 825
           : :| | | || :|  |  : :: | |    | ::| | :  |:  | |:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CE 827
          |:
Qy    161 CD 162


RESULT    6
```

```
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   SLIT PROTEIN PRECURSOR.
GN   SLI.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91099665.
RA   ROTHBERG J.M., JACOBS J.R., GOODMAN C.S., ARTAVANIS-TSAKONAS S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   GENES DEV. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC       BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC       AND TWO BLOCKS OF 5 LRR'S.
CC   -!- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; G8615; -.
DR   PIR; A36665; A36665.
DR   FLYBASE; FBgn0003425; sli.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00008; EGF; 7.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 10.
DR   HSSP; P00740; 1IXA.
KW   NEUROGENESIS; GLYCOPROTEIN; SIGNAL; ALTERNATIVE SPLICING;
KW   EGF-LIKE DOMAIN; REPEAT; LEUCINE-REPEAT; DUPLICATION.
FT   SIGNAL       1     36
FT   CHAIN       37   1480       SLIT PROTEIN.
FT   DOMAIN      70    104       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     105    230       LEUCINE-RICH REPEATS (1ST REGION).
FT   DOMAIN     231    294       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     295    326       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     327    452       LEUCINE-RICH REPEATS (2ND REGION).
FT   DOMAIN     453    518       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     519    550       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     551    653       LEUCINE-RICH REPEATS (3RD REGION).
FT   DOMAIN     654    714       CONSERVED C-FLANKING REGION OF THE LRR.
FT   DOMAIN     715    746       CONSERVED N-FLANKING REGION OF THE LRR.
FT   DOMAIN     747    848       LEUCINE-RICH REPEATS (4TH REGION).
FT   DOMAIN     849    910       CONSERVED C-FLANKING REGION OF THE LRR.
FT   REPEAT     105    115       LRR 1-1.
FT   REPEAT     116    139       LRR 1-2.
FT   REPEAT     140    163       LRR 1-3.
FT   REPEAT     164    187       LRR 1-4.
FT   REPEAT     188    211       LRR 1-5.
FT   REPEAT     212    230       LRR 1-6.
FT   REPEAT     327    337       LRR 2-1.
FT   REPEAT     338    361       LRR 2-2.
FT   REPEAT     362    385       LRR 2-3.
FT   REPEAT     386    409       LRR 2-4.
FT   REPEAT     410    433       LRR 2-5.
FT   REPEAT     434    452       LRR 2-6.
FT   REPEAT     551    562       LRR 3-1.
FT   REPEAT     563    586       LRR 3-2.
FT   REPEAT     587    610       LRR 3-3.
FT   REPEAT     611    634       LRR 3-4.
FT   REPEAT     635    653       LRR 3-5.
FT   REPEAT     747    757       LRR 4-1.
FT   REPEAT     758    781       LRR 4-2.
FT   REPEAT     782    805       LRR 4-3.
FT   REPEAT     806    829       LRR 4-4.
FT   REPEAT     830    848       LRR 4-5.
FT   DOMAIN     907    944       EGF-LIKE 1.
FT   DOMAIN     946    983       EGF-LIKE 2.
FT   DOMAIN     985   1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024   1062       EGF-LIKE 4.
FT   DOMAIN    1064   1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111   1149       EGF-LIKE 6.
FT   DOMAIN    1353   1392       EGF-LIKE 7.
FT   DOMAIN    1409   1480       CTCK.
FT   VARSPLIC  1394   1404       MISSING (IN SHORT FORM).
FT   CARBOHYD   111    111       POTENTIAL.
FT   CARBOHYD   207    207       POTENTIAL.
FT   CARBOHYD   357    357       POTENTIAL.
FT   CARBOHYD   435    435       POTENTIAL.
FT   CARBOHYD   783    783       POTENTIAL.
FT   CARBOHYD   788    788       POTENTIAL.
FT   CARBOHYD   958    958       POTENTIAL.
FT   CARBOHYD   998    998       POTENTIAL.
FT   CARBOHYD  1060   1060       POTENTIAL.
FT   CARBOHYD  1159   1159       POTENTIAL.
FT   CARBOHYD  1175   1175       POTENTIAL.
FT   CARBOHYD  1243   1243       POTENTIAL.
FT   CARBOHYD  1292   1292       POTENTIAL.
FT   DISULFID   911    922       BY SIMILARITY.
FT   DISULFID   916    932       BY SIMILARITY.
FT   DISULFID   934    943       BY SIMILARITY.
FT   DISULFID   950    961       BY SIMILARITY.
FT   DISULFID   955    971       BY SIMILARITY.
FT   DISULFID   973    982       BY SIMILARITY.
FT   DISULFID   989   1001       BY SIMILARITY.
FT   DISULFID   995   1010       BY SIMILARITY.
FT   DISULFID  1012   1021       BY SIMILARITY.
FT   DISULFID  1028   1041       BY SIMILARITY.
FT   DISULFID  1035   1050       BY SIMILARITY.
FT   DISULFID  1052   1061       BY SIMILARITY.
FT   DISULFID  1068   1079       BY SIMILARITY.
FT   DISULFID  1073   1088       BY SIMILARITY.
FT   DISULFID  1090   1099       BY SIMILARITY.
FT   DISULFID  1115   1125       BY SIMILARITY.
FT   DISULFID  1120   1137       BY SIMILARITY.
FT   DISULFID  1139   1148       BY SIMILARITY.
FT   DISULFID  1357   1368       BY SIMILARITY.
FT   DISULFID  1362   1380       BY SIMILARITY.
FT   DISULFID  1382   1391       BY SIMILARITY.
FT   DISULFID  1409   1443       BY SIMILARITY.
FT   DISULFID  1423   1457       BY SIMILARITY.
FT   DISULFID  1434   1473       BY SIMILARITY.
FT   DISULFID  1438   1475       BY SIMILARITY.
FT   DISULFID  1442   1479       BY SIMILARITY.
SQ   SEQUENCE   1480 AA;  165752 MW;  2CD1C421 CRC32;

  Query Match          13.3%;  Score 254;  DB 1;  Length 1480;
  Best Local Similarity  32.6%;  Pred. No. 3.86e-38;
  Matches   45;  Conservative   30;  Mismatches  51;  Indels  12;  Gaps  11;
```

```
Db    900 VRNDILAKCNACFEQPCQNQAQCVALPQREYQCLCQPGYHGKHCEFMI-DACYGNPCRNN 958
          ::  ||: |::|    |  :: | : :|  : | |: |  |  |:    |:| || |  :
Qy     39 MQTGILPGCEPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKC-VH 96

Db    959 ATCTVLEEGRFSCQCAPGYTGARCETNID--D-CLGEIKCQNNATC-IDGV-ESYKCECQ 1013
          :||  ::   :|| |  |  |  |: : |  : |   ||| ::: | : || ::| |||
Qy     97 GTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQ-MIKC-KHGKCRLSGVGQPY-CECN 153

Db   1014 PGFSGEFCDTKIQFCSPE 1031
          :||:|: ||  |  |  |
Qy    154 SGFTGDSCDREIS-CRGE 170


RESULT    7
ID    NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC    P46530;
[illegible]  01-NOV-1995 (REL. 32, CREATED)
      01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
[illegible]  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE    NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN    NOTCH.
OS    BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC    TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC    CYPRINIDAE; RASBORINAE; DANIO.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=EMBRYO;
RX    MEDLINE; 94128602.
RA    BIERKAMP C., CAMPOS-ORTEGA J.A.;
RT    "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT    its pattern of transcription during early embryogenesis.";
RL    MECH. DEV. 43:87-100(1993).
CC    -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC        EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC        NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC    -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC    -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC        STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC        ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC        NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC        ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC        AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC        BRAIN AND HEAD REGIONS.
[illegible]  -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
[illegible]  -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
[illegible]  -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC    -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; X69088; G433867; -.
DR    PROSITE; PS00010; ASX_HYDROXYL; 23.
DR    PROSITE; PS00022; EGF_1; 34.
DR    PROSITE; PS01186; EGF_2; 28.
DR    PROSITE; PS01187; EGF_CA; 22.
DR    PFAM; PF00008; EGF; 36.
DR    PFAM; PF00023; ank; 6.
DR    PFAM; PF00066; notch; 3.
DR    HSSP; P00740; 1IXA.
KW    DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW    TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT    SIGNAL       1     20       POTENTIAL.
FT    CHAIN       21   2437       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT    DOMAIN      21   1724       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM  1725   1747       POTENTIAL.
FT    DOMAIN    1748   2437       CYTOPLASMIC (POTENTIAL).
FT    DOMAIN      21     57       EGF-LIKE 1.
FT    DOMAIN      58     98       EGF-LIKE 2.
FT    DOMAIN     101    138       EGF-LIKE 3.
FT    DOMAIN     139    175       EGF-LIKE 4.
FT    DOMAIN     177    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     217    254       EGF-LIKE 6.
FT    DOMAIN     256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     294    332       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     334    370       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     371    409       EGF-LIKE 10.
FT    DOMAIN     411    449       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     451    487       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     489    524       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     526    562       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     564    599       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     601    637       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     639    674       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     676    712       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     714    749       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     751    787       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     789    825       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     827    865       EGF-LIKE 22.
FT    DOMAIN     867    903       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     905    941       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     943    979       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN     981   1017       EGF-LIKE 26.
FT    DOMAIN    1019   1055       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN    1057   1093       EGF-LIKE 28.
FT    DOMAIN    1095   1141       EGF-LIKE 29.
FT    DOMAIN    1143   1179       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN    1181   1217       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN    1219   1263       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT    DOMAIN    1265   1303       EGF-LIKE 33.
FT    DOMAIN    1305   1344       EGF-LIKE 34.
FT    DOMAIN    1346   1382       EGF-LIKE 35.
FT    DOMAIN    1385   1423       EGF-LIKE 36.
FT    DOMAIN    1446   1561       3 X LIN/NOTCH REPEATS.
FT    REPEAT    1446   1486       LIN/NOTCH 1.
FT    REPEAT    1487   1520       LIN/NOTCH 2.
FT    REPEAT    1521   1561       LIN/NOTCH 3.
FT    DOMAIN    1861   2074       6 X ANK MOTIF REPEATS.
FT    REPEAT    1861   1891       ANK MOTIF 1.
FT    REPEAT    1892   1940       ANK MOTIF 1.
FT    REPEAT    1941   1974       ANK MOTIF 1.
FT    REPEAT    1975   2007       ANK MOTIF 1.
FT    REPEAT    2008   2040       ANK MOTIF 1.
FT    REPEAT    2041   2074       ANK MOTIF 1.
FT    DOMAIN    2265   2276       POLY-GLN (OPA-REPEAT).
FT    DISULFID    25     35       BY SIMILARITY.
FT    DISULFID    29     45       BY SIMILARITY.
FT    DISULFID    47     56       BY SIMILARITY.
FT    DISULFID    62     73       BY SIMILARITY.
FT    DISULFID    67     86       BY SIMILARITY.
FT    DISULFID    88     97       BY SIMILARITY.
FT    DISULFID   105    116       BY SIMILARITY.
FT    DISULFID   110    126       BY SIMILARITY.
FT    DISULFID   128    137       BY SIMILARITY.
FT    DISULFID   143    154       BY SIMILARITY.
FT    DISULFID   148    163       BY SIMILARITY.
FT    DISULFID   165    174       BY SIMILARITY.
FT    DISULFID   181    194       BY SIMILARITY.
FT    DISULFID   188    203       BY SIMILARITY.
FT    DISULFID   205    214       BY SIMILARITY.
FT    DISULFID   221    232       BY SIMILARITY.
FT    DISULFID   226    242       BY SIMILARITY.
FT    DISULFID   244    253       BY SIMILARITY.
FT    DISULFID   260    271       BY SIMILARITY.
FT    DISULFID   265    280       BY SIMILARITY.
FT    DISULFID   282    291       BY SIMILARITY.
FT    DISULFID   298    311       BY SIMILARITY.
FT    DISULFID   305    320       BY SIMILARITY.
FT    DISULFID   322    331       BY SIMILARITY.
FT    DISULFID   338    349       BY SIMILARITY.
```

```
FT   DISULFID    343    358       BY SIMILARITY.
FT   DISULFID    360    369       BY SIMILARITY.
FT   DISULFID    375    386       BY SIMILARITY.
FT   DISULFID    380    397       BY SIMILARITY.
FT   DISULFID    399    408       BY SIMILARITY.
FT   DISULFID    415    428       BY SIMILARITY.
FT   DISULFID    422    437       BY SIMILARITY.
FT   DISULFID    439    448       BY SIMILARITY.
FT   DISULFID    455    466       BY SIMILARITY.
FT   DISULFID    460    475       BY SIMILARITY.
FT   DISULFID    477    486       BY SIMILARITY.
FT   DISULFID    493    503       BY SIMILARITY.
FT   DISULFID    498    512       BY SIMILARITY.
FT   DISULFID    514    523       BY SIMILARITY.
FT   DISULFID    530    541       BY SIMILARITY.
FT   DISULFID    535    550       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    568    578       BY SIMILARITY.
FT   DISULFID    573    587       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    605    616       BY SIMILARITY.
FT   DISULFID    610    625       BY SIMILARITY.
FT   DISULFID    627    636       BY SIMILARITY.
FT   DISULFID    643    653       BY SIMILARITY.
FT   DISULFID    648    662       BY SIMILARITY.
FT   DISULFID    664    673       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   DISULFID    685    700       BY SIMILARITY.
FT   DISULFID    702    711       BY SIMILARITY.
FT   DISULFID    718    728       BY SIMILARITY.
FT   DISULFID    723    737       BY SIMILARITY.
FT   DISULFID    739    748       BY SIMILARITY.
FT   DISULFID    755    766       BY SIMILARITY.
FT   DISULFID    760    775       BY SIMILARITY.
FT   DISULFID    777    786       BY SIMILARITY.
FT   DISULFID    793    804       BY SIMILARITY.
FT   DISULFID    798    813       BY SIMILARITY.
FT   DISULFID    815    824       BY SIMILARITY.
FT   DISULFID    831    842       BY SIMILARITY.
FT   DISULFID    836    853       BY SIMILARITY.
FT   DISULFID    855    864       BY SIMILARITY.
FT   DISULFID    871    882       BY SIMILARITY.
FT   DISULFID    876    891       BY SIMILARITY.
FT   DISULFID    893    902       BY SIMILARITY.
FT   DISULFID    909    920       BY SIMILARITY.
FT   DISULFID    914    929       BY SIMILARITY.
FT   DISULFID    931    940       BY SIMILARITY.
FT   DISULFID    947    958       BY SIMILARITY.
FT   DISULFID    952    967       BY SIMILARITY.
FT   DISULFID    969    978       BY SIMILARITY.
FT   DISULFID   1023   1034       BY SIMILARITY.
FT   DISULFID   1028   1043       BY SIMILARITY.
FT   DISULFID   1045   1054       BY SIMILARITY.
FT   DISULFID   1061   1072       BY SIMILARITY.
FT   DISULFID   1066   1081       BY SIMILARITY.
FT   DISULFID   1083   1092       BY SIMILARITY.
FT   DISULFID   1099   1120       BY SIMILARITY.
FT   DISULFID   1114   1129       BY SIMILARITY.
FT   DISULFID   1131   1140       BY SIMILARITY.
FT   DISULFID   1147   1158       BY SIMILARITY.
FT   DISULFID   1152   1167       BY SIMILARITY.
FT   DISULFID   1169   1178       BY SIMILARITY.
FT   DISULFID   1185   1196       BY SIMILARITY.
FT   DISULFID   1190   1205       BY SIMILARITY.
FT   DISULFID   1207   1216       BY SIMILARITY.
FT   DISULFID   1223   1242       BY SIMILARITY.
FT   DISULFID   1236   1251       BY SIMILARITY.
FT   DISULFID   1253   1262       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          13.2%;  Score 251;  DB 1;  Length 2437;
  Best Local Similarity  34.4%;  Pred. No. 2.16e-37;

    Matches   42;  Conservative   23;  Mismatches  45;  Indels  12;  Gaps  11;

Db    718 CSSNPCIHGSCLDQINS-YRCVCEAGWMGRNCDININE-CLSNPCVNGGTCKDMTS-GYL 774
          |  : | || |     | : | || |||| ||   |: ||:| ||:| ||  : : :|
Qy     50 CHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYS 108

Db    775 CTCRAGFSGPNC--QMNI-NECASNPCLNQGSC-IDDVA-GFKCNCMLPYTGEVCENVLA 829
          | |  | :|  |  : :: | |    | ::| | :  |:  : |:|   :||: |:  ::
Qy    109 CKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDSCDREIS 166

Db    830 PC 831
          |
Qy    167 -C 167


RESULT    8
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   TEPASS U., THERES C., KNUST E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   CELL 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA   VAESSIN H., CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33753; G552087; ALT_SEQ.
DR   EMBL; X05144; E1746; -.
DR   EMBL; X05144; G929536; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   FLYBASE; FBgn0000368; crb.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00054; laminin_G; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
```

```
KW   GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
FT   SIGNAL       1    90
FT   CHAIN       91  2139       CRUMBS PROTEIN.
FT   DOMAIN      91  2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  2085  2111       POTENTIAL.
FT   DOMAIN    2112  2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     267   303       EGF-LIKE 1.
FT   DOMAIN     306   343       EGF-LIKE 2.
FT   DOMAIN     348   386       EGF-LIKE 3.
FT   DOMAIN     388   425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     427   463       EGF-LIKE 5.
FT   DOMAIN     464   500       EGF-LIKE 6.
FT   DOMAIN     501   532       EGF-LIKE 7.
FT   DOMAIN     545   581       EGF-LIKE 8.
FT   DOMAIN     582   611       EGF-LIKE 9.
FT   DOMAIN     609   646       EGF-LIKE 10.
FT   DOMAIN     648   685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     687   723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     725   761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     763   800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     802   838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     840   902       EGF-LIKE 16.
FT   DOMAIN     904   940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     942   978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     980  1021       EGF-LIKE 19.
FT   DOMAIN    1207  1243       EGF-LIKE 20.
FT   DOMAIN    1481  1517       EGF-LIKE 21.
FT   DOMAIN    1759  1795       EGF-LIKE 22.
FT   DOMAIN    1797  1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1835  1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1874  1915       EGF-LIKE 25.
FT   DOMAIN    1915  1951       EGF-LIKE 26.
FT   DOMAIN    1953  1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1991  2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    2030  2070       EGF-LIKE 29.
FT   DISULFID   271   282       BY SIMILARITY.
FT   DISULFID   276   291       BY SIMILARITY.
FT   DISULFID   293   302       BY SIMILARITY.
FT   DISULFID   310   321       BY SIMILARITY.
FT   DISULFID   315   331       BY SIMILARITY.
FT   DISULFID   333   342       BY SIMILARITY.
FT   DISULFID   352   363       BY SIMILARITY.
FT   DISULFID   357   374       BY SIMILARITY.
FT   DISULFID   376   385       BY SIMILARITY.
FT   DISULFID   392   403       BY SIMILARITY.
FT   DISULFID   397   412       BY SIMILARITY.
FT   DISULFID   414   424       BY SIMILARITY.
FT   DISULFID   431   442       BY SIMILARITY.
FT   DISULFID   436   451       BY SIMILARITY.
FT   DISULFID   453   462       BY SIMILARITY.
FT   DISULFID   468   479       BY SIMILARITY.
FT   DISULFID   473   488       BY SIMILARITY.
FT   DISULFID   490   499       BY SIMILARITY.
FT   DISULFID   505   515       BY SIMILARITY.
FT   DISULFID   509   520       BY SIMILARITY.
FT   DISULFID   522   531       BY SIMILARITY.
FT   DISULFID   549   562       BY SIMILARITY.
FT   DISULFID   556   569       BY SIMILARITY.
FT   DISULFID   571   580       BY SIMILARITY.
FT   DISULFID   586   597       BY SIMILARITY.
FT   DISULFID   591   602       BY SIMILARITY.
FT   DISULFID   604   610       BY SIMILARITY.
FT   DISULFID   613   624       BY SIMILARITY.
FT   DISULFID   618   634       BY SIMILARITY.
FT   DISULFID   636   645       BY SIMILARITY.
FT   DISULFID   652   664       BY SIMILARITY.
FT   DISULFID   659   673       BY SIMILARITY.
FT   DISULFID   675   684       BY SIMILARITY.
FT   DISULFID   691   702       BY SIMILARITY.
FT   DISULFID   696   711       BY SIMILARITY.
FT   DISULFID   713   722       BY SIMILARITY.
FT   DISULFID   729   740       BY SIMILARITY.
FT   DISULFID   734   749       BY SIMILARITY.
FT   DISULFID   751   760       BY SIMILARITY.
FT   DISULFID   767   778       BY SIMILARITY.
FT   DISULFID   772   787       BY SIMILARITY.
FT   DISULFID   789   799       BY SIMILARITY.
FT   DISULFID   806   817       BY SIMILARITY.
FT   DISULFID   811   826       BY SIMILARITY.
FT   DISULFID   828   837       BY SIMILARITY.
FT   DISULFID   844   855       BY SIMILARITY.
FT   DISULFID   849   890       BY SIMILARITY.
FT   DISULFID   892   901       BY SIMILARITY.
FT   DISULFID   908   919       BY SIMILARITY.
FT   DISULFID   913   928       BY SIMILARITY.
FT   DISULFID   930   939       BY SIMILARITY.
FT   DISULFID   946   957       BY SIMILARITY.
FT   DISULFID   952   966       BY SIMILARITY.
FT   DISULFID   968   977       BY SIMILARITY.
FT   DISULFID   984   995       BY SIMILARITY.
FT   DISULFID   989  1009       BY SIMILARITY.
FT   DISULFID  1011  1020       BY SIMILARITY.
FT   DISULFID  1211  1222       BY SIMILARITY.
FT   DISULFID  1216  1231       BY SIMILARITY.
FT   DISULFID  1233  1242       BY SIMILARITY.
FT   DISULFID  1485  1496       BY SIMILARITY.
FT   DISULFID  1490  1505       BY SIMILARITY.
FT   DISULFID  1507  1516       BY SIMILARITY.
FT   DISULFID  1763  1774       BY SIMILARITY.
FT   DISULFID  1768  1783       BY SIMILARITY.
FT   DISULFID  1785  1794       BY SIMILARITY.
FT   DISULFID  1801  1812       BY SIMILARITY.
FT   DISULFID  1806  1821       BY SIMILARITY.
FT   DISULFID  1823  1832       BY SIMILARITY.
FT   DISULFID  1839  1850       BY SIMILARITY.
FT   DISULFID  1844  1859       BY SIMILARITY.
FT   DISULFID  1861  1870       BY SIMILARITY.
FT   DISULFID  1878  1889       BY SIMILARITY.
FT   DISULFID  1883  1903       BY SIMILARITY.
FT   DISULFID  1905  1914       BY SIMILARITY.
FT   DISULFID  1919  1930       BY SIMILARITY.
FT   DISULFID  1924  1939       BY SIMILARITY.
FT   DISULFID  1941  1950       BY SIMILARITY.
FT   DISULFID  1957  1968       BY SIMILARITY.
FT   DISULFID  1962  1977       BY SIMILARITY.
FT   DISULFID  1979  1988       BY SIMILARITY.
FT   DISULFID  1995  2008       BY SIMILARITY.
FT   DISULFID  2002  2017       BY SIMILARITY.
FT   DISULFID  2019  2028       BY SIMILARITY.
FT   CARBOHYD    37    37       POTENTIAL.
FT   CARBOHYD    96    96       POTENTIAL.
FT   CARBOHYD   198   198       POTENTIAL.
FT   CARBOHYD   238   238       POTENTIAL.
FT   CARBOHYD   239   239       POTENTIAL.
FT   CARBOHYD   336   336       POTENTIAL.
FT   CARBOHYD   400   400       POTENTIAL.
FT   CARBOHYD   550   550       POTENTIAL.
FT   CARBOHYD   565   565       POTENTIAL.
FT   CARBOHYD   736   736       POTENTIAL.
FT   CARBOHYD   746   746       POTENTIAL.
FT   CARBOHYD   860   860       POTENTIAL.
FT   CARBOHYD   884   884       POTENTIAL.
FT   CARBOHYD   976   976       POTENTIAL.
FT   CARBOHYD  1102  1102       POTENTIAL.
FT   CARBOHYD  1114  1114       POTENTIAL.
FT   CARBOHYD  1138  1138       POTENTIAL.
FT   CARBOHYD  1192  1192       POTENTIAL.
FT   CARBOHYD  1245  1245       POTENTIAL.
FT   CARBOHYD  1255  1255       POTENTIAL.
FT   CARBOHYD  1354  1354       POTENTIAL.
FT   CARBOHYD  1363  1363       POTENTIAL.
FT   CARBOHYD  1441  1441       POTENTIAL.
FT   CARBOHYD  1454  1454       POTENTIAL.
...
Note: remainder of annotations omitted.
```

```
  Query Match          12.9%;  Score 245;  DB 1;  Length 2139;
  Best Local Similarity  31.7%;  Pred. No. 6.68e-36;
  Matches   40;  Conservative   23;  Mismatches  57;  Indels   6;  Gaps   6;

Db    271 CLNDPCMGHGTC-SSSPEGYECRCTARYSGKNCQKDNGSPCAKNPCENGGSCLENSEGNY 329
          | :  | :|| | :||  |: | |     |  |:     || | | :| :||  :  :|
Qy     50 CHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSY 107

Db    330 QCFCDPNHSGQHCETEVNIHPLCQTNPCLNNGACVVIGGSGALTCECPKGYAGARCEVDT 389
          | |   |:|  |: | ::   ||   | ::| |  ::| |   |||  |::|  |: :
Qy    108 SCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCR-LSGVGQPYCECNSGFTGDSCDREI 165

Db    390 DECASQ 395
            | ::
Qy    166 S-CRGE 170


RESULT    9
ID   NTC1_HUMAN     STANDARD;      PRT;  2444 AA.
AC   P46531;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA   SMITH S.D., SKLAR J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   CELL 66:649-661(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M73980; G338675; -.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19  >2444       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN      19   1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1737   1757       POTENTIAL.
FT   DOMAIN    1758  >2444       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      20     58       EGF-LIKE 1.
FT   DOMAIN      59     99       EGF-LIKE 2.
FT   DOMAIN     102    139       EGF-LIKE 3.
FT   DOMAIN     140    176       EGF-LIKE 4.
FT   DOMAIN     178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     218    255       EGF-LIKE 6.
FT   DOMAIN     257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    410       EGF-LIKE 10.
FT   DOMAIN     412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 20.
FT   DOMAIN     791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    868       EGF-LIKE 22.
FT   DOMAIN     870    906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     908    944       EGF-LIKE 24.
FT   DOMAIN     946    982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 26.
FT   DOMAIN    1022   1058       EGF-LIKE 27.
FT   DOMAIN    1060   1096       EGF-LIKE 28.
FT   DOMAIN    1098   1144       EGF-LIKE 29.
FT   DOMAIN    1146   1182       EGF-LIKE 30.
FT   DOMAIN    1184   1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1222   1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1268   1306       EGF-LIKE 33.
FT   DOMAIN    1308   1347       EGF-LIKE 34.
FT   DOMAIN    1349   1385       EGF-LIKE 35.
FT   DOMAIN    1388   1427       EGF-LIKE 36.
FT   DOMAIN    1446   1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1446   1481       LIN/NOTCH 1.
FT   REPEAT    1482   1523       LIN/NOTCH 2.
FT   REPEAT    1524   1563       LIN/NOTCH 3.
FT   DOMAIN    1876   2087       6 X ANK MOTIF REPEATS.
FT   REPEAT    1876   1921       ANK MOTIF 1.
FT   REPEAT    1923   1954       ANK MOTIF 2.
FT   REPEAT    1956   1987       ANK MOTIF 3.
FT   REPEAT    1990   2021       ANK MOTIF 4.
FT   REPEAT    2023   2054       ANK MOTIF 5.
FT   REPEAT    2056   2087       ANK MOTIF 6.
FT   DOMAIN    1576   1579       POLY-VAL.
FT   DOMAIN    1662   1665       POLY-ARG.
FT   DOMAIN    1729   1732       POLY-PRO.
FT   DOMAIN    1741   1744       POLY-ALA.
FT   DOMAIN    1902   1905       POLY-GLU.
FT   DOMAIN    2260   2263       POLY-GLY.
FT   DOMAIN    2404   2407       POLY-GLN.
FT   DOMAIN    2411   2418       POLY-PRO.
FT   DISULFID    24     37       BY SIMILARITY.
FT   DISULFID    31     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    68     87       BY SIMILARITY.
FT   DISULFID    89     98       BY SIMILARITY.
FT   DISULFID   106    117       BY SIMILARITY.
FT   DISULFID   111    127       BY SIMILARITY.
FT   DISULFID   129    138       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   182    195       BY SIMILARITY.
FT   DISULFID   189    204       BY SIMILARITY.
FT   DISULFID   206    215       BY SIMILARITY.
FT   DISULFID   222    233       BY SIMILARITY.
FT   DISULFID   227    243       BY SIMILARITY.
```

```
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    261    272       BY SIMILARITY.
FT   DISULFID    266    281       BY SIMILARITY.
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    299    312       BY SIMILARITY.
FT   DISULFID    306    321       BY SIMILARITY.
FT   DISULFID    323    332       BY SIMILARITY.
FT   DISULFID    339    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    361    370       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    398       BY SIMILARITY.
FT   DISULFID    400    409       BY SIMILARITY.
FT   DISULFID    416    429       BY SIMILARITY.
FT   DISULFID    423    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    456    467       BY SIMILARITY.
FT   DISULFID    461    476       BY SIMILARITY.
FT   DISULFID    478    487       BY SIMILARITY.
FT   DISULFID    494    505       BY SIMILARITY.
FT   DISULFID    499    514       BY SIMILARITY.
FT   DISULFID    516    525       BY SIMILARITY.
FT   DISULFID    532    543       BY SIMILARITY.
FT   DISULFID    537    552       BY SIMILARITY.
FT   DISULFID    554    563       BY SIMILARITY.
FT   DISULFID    570    580       BY SIMILARITY.
FT   DISULFID    575    589       BY SIMILARITY.
FT   DISULFID    591    600       BY SIMILARITY.
FT   DISULFID    607    618       BY SIMILARITY.
FT   DISULFID    612    627       BY SIMILARITY.
FT   DISULFID    629    638       BY SIMILARITY.
FT   DISULFID    645    655       BY SIMILARITY.
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    867       BY SIMILARITY.
FT   DISULFID    874    885       BY SIMILARITY.
FT   DISULFID    879    894       BY SIMILARITY.
FT   DISULFID    896    905       BY SIMILARITY.
FT   DISULFID    912    923       BY SIMILARITY.
FT   DISULFID    917    932       BY SIMILARITY.
FT   DISULFID    934    943       BY SIMILARITY.
FT   DISULFID    988    999       BY SIMILARITY.
FT   DISULFID    993   1008       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1026   1037       BY SIMILARITY.
FT   DISULFID   1031   1046       BY SIMILARITY.
FT   DISULFID   1048   1057       BY SIMILARITY.
FT   DISULFID   1064   1075       BY SIMILARITY.
FT   DISULFID   1069   1084       BY SIMILARITY.
FT   DISULFID   1086   1095       BY SIMILARITY.
FT   DISULFID   1102   1123       BY SIMILARITY.
FT   DISULFID   1117   1132       BY SIMILARITY.
FT   DISULFID   1134   1143       BY SIMILARITY.
FT   DISULFID   1150   1161       BY SIMILARITY.
FT   DISULFID   1155   1170       BY SIMILARITY.
FT   DISULFID   1172   1181       BY SIMILARITY.
FT   DISULFID   1188   1199       BY SIMILARITY.
FT   DISULFID   1193   1208       BY SIMILARITY.
...
```

```
Note: remainder of annotations omitted.

  Query Match          12.9%;  Score 246;  DB 1;  Length 2444;
  Best Local Similarity  33.6%;  Pred. No. 3.78e-36;
  Matches   41;  Conservative   26;  Mismatches  44;  Indels  11;  Gaps  10;

Db    714 LSEVNECNSNPCVHGACR-DSLNGYKCDCDPGWSGTNCDINNNE-CESNPCVNGGTCKDM 771
          |:  : |: : | || |:  | :|: |:|: || |  ||   |: | :| ||:| ||  :
Qy     44 LPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPI 102

Db    772 TS-GIVCTCREGFSGPNC--QTNI-NECASNPCLNKGTC-IDDVA-GYKCNCLLPYTGAT 825
           : :  | | || :|  |  : :: | |    | : | | :  |:  | |:|   :|| :
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPY-CECNSGFTGDS 160

Db    826 CE 827
          |:
Qy    161 CD 162


RESULT  10
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; P04154;
DT   01-NOV-1986 (REL. 03, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN   N.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86079539.
RA   WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   CELL 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 87064624.
RA   KIDD S., KELLEY M.R., YOUNG M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   MOL. CELL. BIOL. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE; 85099329.
RA   WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   CELL 40:55-62(1985).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 87257846.
RA   KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   MOL. CELL. BIOL. 7:1545-1548(1987).
RN   [5]
RP   REVIEW.
RA   HARRIS W.A.;
RT   "Many cell types specified by Notch function.";
RL   CURR. BIOL. 1:120-122(1991).
CC   -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC       ECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
```

```
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16152; G157988; -.
DR   EMBL; M16153; G157988; JOINED.
DR   EMBL; M16149; G157988; JOINED.
DR   EMBL; M16150; G157988; JOINED.
DR   EMBL; M16151; G157988; JOINED.
DR   EMBL; K03508; G157993; -.
DR   EMBL; M13689; G157993; JOINED.
DR   EMBL; K03507; G157993; JOINED.
DR   EMBL; M12175; G950317; -.
DR   EMBL; M16025; G157995; -.
DR   PIR; A24420; A24420.
DR   PIR; A24768; A24768.
DR   PIR; A05267; A05267.
DR   FLYBASE; FBgn0004647; N.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     44       POTENTIAL.
FT   CHAIN       45   2703       NEUROGENIC LOCUS NOTCH PROTEIN.
FT   DOMAIN      45   1745       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1746   1766       POTENTIAL.
FT   DOMAIN    1767   2703       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58   1451       36 X EGF-TYPE REPEATS.
FT   DOMAIN      58     95       EGF-LIKE 1.
FT   DOMAIN      96    136       EGF-LIKE 2.
FT   DOMAIN     139    176       EGF-LIKE 3.
FT   DOMAIN     177    215       EGF-LIKE 4.
FT   DOMAIN     217    253       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     255    291       EGF-LIKE 6.
FT   DOMAIN     293    329       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     331    370       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     372    408       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     409    447       EGF-LIKE 10.
FT   DOMAIN     449    486       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     488    524       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     526    562       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     564    600       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     602    637       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     639    675       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     677    713       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     715    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     829    865       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     867    905       EGF-LIKE 22.
FT   DOMAIN     907    944       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     946    982       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     984   1020       EGF-LIKE 25.
FT   DOMAIN    1022   1058       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1060   1096       EGF-LIKE 27.
FT   DOMAIN    1098   1134       EGF-LIKE 28.
FT   DOMAIN    1136   1181       EGF-LIKE 29.
FT   DOMAIN    1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1221   1257       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1259   1295       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1297   1335       EGF-LIKE 33.
FT   DOMAIN    1337   1373       EGF-LIKE 34.
FT   DOMAIN    1375   1412       EGF-LIKE 35.
FT   DOMAIN    1415   1451       EGF-LIKE 36.
FT   DOMAIN    1475   1593       3 X LIN/NOTCH REPEATS.
FT   REPEAT    1475   1513       LIN/NOTCH 1.
FT   REPEAT    1514   1553       LIN/NOTCH 2.
FT   REPEAT    1554   1593       LIN/NOTCH 3.
FT   DOMAIN    1896   2109       6 X ANK MOTIF REPEATS.
FT   DOMAIN    2538   2568       POLY-GLN (OPA-REPEAT).
FT   DISULFID    62     73       BY SIMILARITY.
FT   DISULFID    67     83       BY SIMILARITY.
FT   DISULFID    85     94       BY SIMILARITY.
FT   DISULFID   100    111       BY SIMILARITY.
FT   DISULFID   105    124       BY SIMILARITY.
FT   DISULFID   126    135       BY SIMILARITY.
FT   DISULFID   143    154       BY SIMILARITY.
FT   DISULFID   148    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   181    192       BY SIMILARITY.
FT   DISULFID   186    203       BY SIMILARITY.
FT   DISULFID   205    214       BY SIMILARITY.
FT   DISULFID   221    232       BY SIMILARITY.
FT   DISULFID   226    241       BY SIMILARITY.
FT   DISULFID   243    252       BY SIMILARITY.
FT   DISULFID   259    270       BY SIMILARITY.
FT   DISULFID   264    279       BY SIMILARITY.
FT   DISULFID   281    290       BY SIMILARITY.
FT   DISULFID   297    308       BY SIMILARITY.
FT   DISULFID   302    317       BY SIMILARITY.
FT   DISULFID   319    328       BY SIMILARITY.
FT   DISULFID   335    349       BY SIMILARITY.
FT   DISULFID   343    358       BY SIMILARITY.
FT   DISULFID   360    369       BY SIMILARITY.
FT   DISULFID   376    387       BY SIMILARITY.
FT   DISULFID   381    396       BY SIMILARITY.
FT   DISULFID   398    407       BY SIMILARITY.
FT   DISULFID   413    424       BY SIMILARITY.
FT   DISULFID   418    435       BY SIMILARITY.
FT   DISULFID   437    446       BY SIMILARITY.
FT   DISULFID   453    465       BY SIMILARITY.
FT   DISULFID   459    474       BY SIMILARITY.
FT   DISULFID   476    485       BY SIMILARITY.
FT   DISULFID   492    503       BY SIMILARITY.
FT   DISULFID   497    512       BY SIMILARITY.
FT   DISULFID   514    523       BY SIMILARITY.
FT   DISULFID   530    541       BY SIMILARITY.
FT   DISULFID   535    550       BY SIMILARITY.
FT   DISULFID   552    561       BY SIMILARITY.
FT   DISULFID   568    579       BY SIMILARITY.
FT   DISULFID   573    588       BY SIMILARITY.
FT   DISULFID   590    599       BY SIMILARITY.
FT   DISULFID   606    616       BY SIMILARITY.
FT   DISULFID   611    625       BY SIMILARITY.
FT   DISULFID   627    636       BY SIMILARITY.
FT   DISULFID   643    654       BY SIMILARITY.
FT   DISULFID   648    663       BY SIMILARITY.
FT   DISULFID   665    674       BY SIMILARITY.
FT   DISULFID   681    692       BY SIMILARITY.
FT   DISULFID   686    701       BY SIMILARITY.
FT   DISULFID   703    712       BY SIMILARITY.
FT   DISULFID   719    730       BY SIMILARITY.
FT   DISULFID   724    739       BY SIMILARITY.
FT   DISULFID   741    750       BY SIMILARITY.
FT   DISULFID   757    768       BY SIMILARITY.
FT   DISULFID   762    777       BY SIMILARITY.
FT   DISULFID   779    788       BY SIMILARITY.
FT   DISULFID   795    806       BY SIMILARITY.
FT   DISULFID   800    815       BY SIMILARITY.
FT   DISULFID   817    826       BY SIMILARITY.
FT   DISULFID   833    844       BY SIMILARITY.
FT   DISULFID   838    853       BY SIMILARITY.
```

```
FT   DISULFID    855    864       BY SIMILARITY.
...
Note: remainder of annotations omitted.

  Query Match          12.8%;  Score 244;  DB 1;  Length 2703;
  Best Local Similarity  31.0%;  Pred. No. 1.18e-35;
  Matches   39;  Conservative   30;  Mismatches  47;  Indels  10;  Gaps  10;

Db    566 DDCQSQPCRNRGICH-DSIAGYSCECPPGYTGTSCEININD-CDSNPCHRGKCID-DVNS 622
          : |:   |  :| |:  | :|::|||  |  |  |:   || | :| | :| |:  :  |
Qy     48 EPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFS 106

Db    623 FKCLCDPGYTGYIC-QKQ-I-NECESNPCQFDGHCQ-DRVGSYYCQCQAGTSGKNCEVNV 678
          : | |  |  | :| : : : | |:   |   | |:   ||  ||:| :|.:| :|: ::
Qy    107 YSCKCLEGHGGVLCDEEEDLFNPCQMIKCK-HGKCRLSGVGQPYCECNSGFTGDSCDREI 165

     679 NECHSN 684
         : |:::
     166 S-CRGE 170


RESULT   11
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
DE   (EGF III) (FIBROPELLIN III).
GN   EGF3.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GASTRULA;
RX   MEDLINE; 93273088.
RA   BISGROVE B.W., RAFF R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   DEV. BIOL. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
     -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
     -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07045; G310660; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PFAM; PF00008; EGF; 8.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P00740; 1IXA.
KW   BIOTIN; EGF-LIKE DOMAIN; REPEAT; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    570       FIBROPELLIN C.
FT   DOMAIN      18     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7.
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    570       AVIDIN-LIKE.
FT   DISULFID    23     34       BY SIMILARITY.
FT   DISULFID    28     43       BY SIMILARITY.
FT   DISULFID    45     54       BY SIMILARITY.
FT   DISULFID   180    191       BY SIMILARITY.
FT   DISULFID   185    200       BY SIMILARITY.
FT   DISULFID   202    211       BY SIMILARITY.
FT   DISULFID   218    229       BY SIMILARITY.
FT   DISULFID   223    238       BY SIMILARITY.
FT   DISULFID   240    249       BY SIMILARITY.
FT   DISULFID   256    267       BY SIMILARITY.
FT   DISULFID   261    276       BY SIMILARITY.
FT   DISULFID   278    287       BY SIMILARITY.
FT   DISULFID   294    305       BY SIMILARITY.
FT   DISULFID   299    314       BY SIMILARITY.
FT   DISULFID   316    325       BY SIMILARITY.
FT   DISULFID   332    343       BY SIMILARITY.
FT   DISULFID   337    352       BY SIMILARITY.
FT   DISULFID   354    363       BY SIMILARITY.
FT   DISULFID   370    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   430    439       BY SIMILARITY.
FT   CARBOHYD    30     30       POTENTIAL.
FT   CARBOHYD   136    136       POTENTIAL.
FT   CARBOHYD   357    357       POTENTIAL.
SQ   SEQUENCE   570 AA;  61116 MW;  265BC4BB CRC32;

  Query Match          12.7%;  Score 242;  DB 1;  Length 570;
  Best Local Similarity  32.0%;  Pred. No. 3.69e-35;
  Matches   40;  Conservative   29;  Mismatches  43;  Indels  13;  Gaps  12;

Db    216 DECASAPCRNGGAC-VDQVNGYTCNCIPGFNGVNCENNINE-CASIPCLNGG-ICVDGIN 272
          : |    | :|  |     :|:||:|  |  |  |:   |: | :  |::|  : :::::
Qy     48 EPCHKKVCAHG-CCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFS 106

Db    273 QFACTCLPGYTGILC--ETDI-NECASSPCQNGGSCT-DAVNR-YTCDCRAGFTGSNCET 327
            ::| || |  |:||  | |: | |    | :|  |   :| : | |:| :|||| :|:
Qy    107 -YSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCDR 163

Db    328 NINEC 332
          :|: |
Qy    164 EIS-C 167


RESULT   12
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING
DE   PROTEIN INT-3).
GN   NOTCH4 OR INT3 OR INT-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92194507.
RA   ROBBINS J., BLONDEL B.J., GALLAHAN D., CALLAHAN R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. VIROL. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   CALLAHAN R.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG, AND TESTIS;
RX   MEDLINE; 96281668.
RA   UYTTENDAELE H., MARAZZI G., WU G., YAN Q., SASSOON D., KITAJEWSKI J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   DEVELOPMENT 122:2251-2259(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; G1714084; -.
DR   EMBL; U43691; G1401160; -.
DR   PIR; A38072; TVMVT3.
DR   MGD; MGI:107471; NOTCH4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
DR   PFAM; PF00008; EGF; 26.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; PROTO-ONCOGENE; ANK REPEAT; SIGNAL.
FT   SIGNAL       1     20      POTENTIAL.
FT   CHAIN       21   1964      NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   DOMAIN      21   1443      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1444   1464      POTENTIAL.
FT   DOMAIN    1465   1964      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21     60      EGF-LIKE 1.
FT   DOMAIN      61    112      EGF-LIKE 2.
FT   DOMAIN     115    152      EGF-LIKE 3.
FT   DOMAIN     153    189      EGF-LIKE 4.
FT   DOMAIN     191    229      EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     231    271      EGF-LIKE 6.
FT   DOMAIN     273    309      EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     311    350      EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    388      EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     389    427      EGF-LIKE 10.
FT   DOMAIN     429    470      EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     472    508      EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     510    546      EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     548    584      EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     586    622      EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     622    656      EGF-LIKE 16.
FT   DOMAIN     658    686      EGF-LIKE 17.
FT   DOMAIN     688    724      EGF-LIKE 18.
FT   DOMAIN     726    762      EGF-LIKE 19.
FT   DOMAIN     764    800      EGF-LIKE 20.
FT   DOMAIN     803    839      EGF-LIKE 21.
FT   DOMAIN     841    877      EGF-LIKE 22.
FT   DOMAIN     878    924      EGF-LIKE 23.
FT   DOMAIN     926    962      EGF-LIKE 24.
FT   DOMAIN     964   1000      EGF-LIKE 25.
FT   DOMAIN    1002   1040      EGF-LIKE 26.
FT   DOMAIN    1042   1081      EGF-LIKE 27.
FT   DOMAIN    1083   1122      EGF-LIKE 28.
FT   DOMAIN    1126   1167      EGF-LIKE 29.
FT   DOMAIN    1168   1282      3 X LIN/NOTCH REPEATS.
FT   REPEAT    1168   1208      LIN/NOTCH 1.
FT   REPEAT    1209   1242      LIN/NOTCH 2.
FT   REPEAT    1243   1282      LIN/NOTCH 3.
FT   DOMAIN    1572   1785      6 X ANK MOTIF REPEATS.
FT   REPEAT    1572   1603      ANK MOTIF 1.
FT   REPEAT    1622   1653      ANK MOTIF 2.
FT   REPEAT    1654   1685      ANK MOTIF 3.
FT   REPEAT    1688   1719      ANK MOTIF 4.
FT   REPEAT    1721   1752      ANK MOTIF 5.
FT   REPEAT    1754   1785      ANK MOTIF 6.
FT   DISULFID    25     38      BY SIMILARITY.
FT   DISULFID    32     48      BY SIMILARITY.
FT   DISULFID    50     59      BY SIMILARITY.
FT   DISULFID    65     77      BY SIMILARITY.
FT   DISULFID    71    100      BY SIMILARITY.
FT   DISULFID   102    111      BY SIMILARITY.
FT   DISULFID   119    130      BY SIMILARITY.
FT   DISULFID   124    140      BY SIMILARITY.
FT   DISULFID   142    151      BY SIMILARITY.
FT   DISULFID   157    168      BY SIMILARITY.
FT   DISULFID   162    177      BY SIMILARITY.
FT   DISULFID   179    188      BY SIMILARITY.
FT   DISULFID   195    208      BY SIMILARITY.
FT   DISULFID   202    217      BY SIMILARITY.
FT   DISULFID   219    228      BY SIMILARITY.
FT   DISULFID   235    246      BY SIMILARITY.
FT   DISULFID   240    259      BY SIMILARITY.
FT   DISULFID   261    270      BY SIMILARITY.
FT   DISULFID   235    246      BY SIMILARITY.
FT   DISULFID   240    259      BY SIMILARITY.
FT   DISULFID   261    270      BY SIMILARITY.
FT   DISULFID   277    288      BY SIMILARITY.
FT   DISULFID   282    297      BY SIMILARITY.
FT   DISULFID   299    308      BY SIMILARITY.
FT   DISULFID   315    329      BY SIMILARITY.
FT   DISULFID   323    338      BY SIMILARITY.
FT   DISULFID   340    349      BY SIMILARITY.
FT   DISULFID   356    367      BY SIMILARITY.
FT   DISULFID   361    376      BY SIMILARITY.
FT   DISULFID   378    387      BY SIMILARITY.
FT   DISULFID   393    404      BY SIMILARITY.
FT   DISULFID   398    415      BY SIMILARITY.
FT   DISULFID   417    426      BY SIMILARITY.
FT   DISULFID   433    449      BY SIMILARITY.
FT   DISULFID   443    458      BY SIMILARITY.
FT   DISULFID   460    469      BY SIMILARITY.
FT   DISULFID   476    487      BY SIMILARITY.
FT   DISULFID   481    496      BY SIMILARITY.
FT   DISULFID   498    507      BY SIMILARITY.
FT   DISULFID   514    525      BY SIMILARITY.
FT   DISULFID   519    534      BY SIMILARITY.
FT   DISULFID   536    545      BY SIMILARITY.
FT   DISULFID   552    563      BY SIMILARITY.
FT   DISULFID   557    572      BY SIMILARITY.
FT   DISULFID   574    583      BY SIMILARITY.
FT   DISULFID   590    601      BY SIMILARITY.
FT   DISULFID   595    610      BY SIMILARITY.
FT   DISULFID   612    621      BY SIMILARITY.
FT   DISULFID   626    637      BY SIMILARITY.
FT   DISULFID   631    646      BY SIMILARITY.
FT   DISULFID   648    655      BY SIMILARITY.
FT   DISULFID   662    669      BY SIMILARITY.
FT   DISULFID   664    674      BY SIMILARITY.
FT   DISULFID   676    685      BY SIMILARITY.
```

```
FT   DISULFID   692   703       BY SIMILARITY.
FT   DISULFID   697   712       BY SIMILARITY.
FT   DISULFID   714   723       BY SIMILARITY.
FT   DISULFID   730   741       BY SIMILARITY.
FT   DISULFID   735   750       BY SIMILARITY.
FT   DISULFID   752   761       BY SIMILARITY.
FT   DISULFID   768   779       BY SIMILARITY.
FT   DISULFID   773   788       BY SIMILARITY.
FT   DISULFID   790   799       BY SIMILARITY.
FT   DISULFID   807   818       BY SIMILARITY.
FT   DISULFID   812   827       BY SIMILARITY.
FT   DISULFID   829   838       BY SIMILARITY.
FT   DISULFID   845   856       BY SIMILARITY.
FT   DISULFID   850   865       BY SIMILARITY.
FT   DISULFID   867   876       BY SIMILARITY.
FT   DISULFID   882   903       BY SIMILARITY.
FT   DISULFID   897   912       BY SIMILARITY.
FT   DISULFID   914   923       BY SIMILARITY.
FT   DISULFID   930   941       BY SIMILARITY.
FT   DISULFID   935   950       BY SIMILARITY.
FT   DISULFID   952   961       BY SIMILARITY.
FT   DISULFID   968   979       BY SIMILARITY.
FT   DISULFID   973   988       BY SIMILARITY.
FT   DISULFID   990   999       BY SIMILARITY.
FT   DISULFID  1006  1019       BY SIMILARITY.
FT   DISULFID  1011  1028       BY SIMILARITY.
FT   DISULFID  1030  1039       BY SIMILARITY.
FT   DISULFID  1046  1057       BY SIMILARITY.
FT   DISULFID  1051  1069       BY SIMILARITY.
FT   DISULFID  1071  1080       BY SIMILARITY.
FT   DISULFID  1087  1098       BY SIMILARITY.
FT   DISULFID  1092  1110       BY SIMILARITY.
FT   DISULFID  1112  1121       BY SIMILARITY.
FT   DISULFID  1130  1142       BY SIMILARITY.
FT   DISULFID  1136  1155       BY SIMILARITY.
FT   DISULFID  1157  1166       BY SIMILARITY.
FT   CARBOHYD   711   711       POTENTIAL.
FT   CARBOHYD   960   960       POTENTIAL.
FT   CARBOHYD  1139  1139       POTENTIAL.
FT   CONFLICT    43    43       Q -> R (IN REF. 3).
FT   CONFLICT   298   298       L -> P (IN REF. 3).
FT   CONFLICT   884   884       M -> K (IN REF. 3).
...
Note: remainder of annotations omitted.

  Query Match          12.7%;  Score 241;  DB 1;  Length 1964;
  Best Local Similarity  37.2%;  Pred. No. 6.52e-35;
  Matches   48;  Conservative   17;  Mismatches  48;  Indels  16;  Gaps  13;

Db    966 EACQSQPCHNHGTC-TSRPGGFHCACPPGFVGLRCEGDVDE-CLDRPCHPSGTAACHSLA 1023
          |:|:   |  || |  |  :|| | |  | :|  |:   :: ||   |   ||  | ::
Qy     48 EPCHKKVC-AHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GT--CLPI- 102

Db   1024 NAF-Y-CQCLPGHTGQRCEVEMDL---CQSQPCSNGGSCEITTGPPPGFTCHCPKGFEGP 1078
          ||| | | || || |  |: | ||   ||   | :|  | ::    | : | |  || |
Qy    103 NAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQP-Y-CECNSGFTGD 159

Db   1079 TCSHKALSC 1087
          :|  : :||
Qy    160 SCD-REISC 167


RESULT   13
ID   GLP1_CAEEL     STANDARD;      PRT;  1295 AA.
AC   P13508;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GLP-1 PROTEIN PRECURSOR.
GN   GLP-1 OR EMB-33 OR F02A9.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 89336787.
RA   YOCHEM J., GREENWALD I.;
RT   "glp-1 and lin-12, genes implicated in distinct cell-cell
RT   interactions in C. elegans, encode similar transmembrane proteins.";
RL   CELL 58:553-563(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [3]
RP   DELETION OF 1174-1295.
RX   MEDLINE; 91351288.
RA   MANGO S.E., MAINE E.M., KIMBLE J.;
RT   "Carboxy-terminal truncation activates glp-1 protein to specify
RT   vulval fates in Caenorhabditis elegans.";
RL   NATURE 352:811-815(1991).
RN   [4]
RP   CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.
RX   MEDLINE; 93354444.
RA   ROEHL H., KIMBLE J.;
RT   "Control of cell fate in C. elegans by a GLP-1 peptide consisting
RT   primarily of ankyrin repeats.";
RL   NATURE 364:632-635(1993).
RN   [5]
RP   FUNCTION.
RX   MEDLINE; 94208066.
RA   MELLO C.C., DRAPER B.W., PRIESS J.R.;
RT   "The maternal genes apx-1 and glp-1 and establishment of
RT   dorsal-ventral polarity in the early C. elegans embryo.";
RL   CELL 77:95-106(1994).
CC   -!- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE
CC       THE BLASTOMERES, ABA AND APA. PROPER SIGNALLING BY GLP-1 INDUCES
CC       ABA DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA
CC       DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE
CC       ESTABLISHMENT THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: ACTS ON ABP DEVELOPMENT DURING 4-CELL AND
CC       12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL
CC       STAGES.
CC   -!- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.
CC   -!- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25580; G156317; -.
DR   EMBL; Z19555; E1322024; -.
DR   EMBL; Z29116; E1322024; JOINED.
DR   EMBL; Z29116; E1323609; -.
```

```
DR   EMBL; Z19555; E1323609; JOINED.
DR   PIR; A32901; A32901.
DR   WORMPEP; F02A9.6; CE00237.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PFAM; PF00008; EGF; 10.
DR   PFAM; PF00023; ank; 4.
DR   PFAM; PF00066; notch; 3.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW   GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1     15       POTENTIAL.
FT   CHAIN       16   1295       GLP-1 PROTEIN.
FT   DOMAIN      16    764       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   765    786       POTENTIAL.
FT   DOMAIN     787   1295       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     493    607       3 X LIN/NOTCH REPEATS.
FT   DOMAIN     988   1133       6 X ANK MOTIF REPEATS.
FT   DOMAIN      19     58       EGF-LIKE 1.
FT   DOMAIN     117    152       EGF-LIKE 2.
FT   DOMAIN     154    190       EGF-LIKE 3.
FT   DOMAIN     190    230       EGF-LIKE 4.
FT   DOMAIN     232    269       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     271    308       EGF-LIKE 6.
FT   DOMAIN     316    359       EGF-LIKE 7.
FT   DOMAIN     369    406       EGF-LIKE 8.
FT   DOMAIN     407    443       EGF-LIKE 9.
FT   DOMAIN     446    479       EGF-LIKE 10.
FT   REPEAT     493    527       LIN/NOTCH 1.
FT   REPEAT     528    568       LIN/NOTCH 2.
FT   REPEAT     569    608       LIN/NOTCH 3.
FT   REPEAT     915    946       ANK MOTIF 1.
FT   REPEAT     947    987       ANK MOTIF 2.
FT   REPEAT     988   1019       ANK MOTIF 3.
FT   REPEAT    1020   1056       ANK MOTIF 4.
FT   REPEAT    1057   1098       ANK MOTIF 5.
FT   REPEAT    1099   1133       ANK MOTIF 6.
FT   DISULFID    23     35       BY SIMILARITY.
FT   DISULFID    29     46       BY SIMILARITY.
FT   DISULFID    48     57       BY SIMILARITY.
FT   DISULFID   121    131       BY SIMILARITY.
FT   DISULFID   126    140       BY SIMILARITY.
FT   DISULFID   142    151       BY SIMILARITY.
FT   DISULFID   158    169       BY SIMILARITY.
FT   DISULFID   163    178       BY SIMILARITY.
FT   DISULFID   180    189       BY SIMILARITY.
FT   DISULFID   194    206       BY SIMILARITY.
FT   DISULFID   201    218       BY SIMILARITY.
FT   DISULFID   220    229       BY SIMILARITY.
FT   DISULFID   236    248       BY SIMILARITY.
FT   DISULFID   242    257       BY SIMILARITY.
FT   DISULFID   259    268       BY SIMILARITY.
FT   DISULFID   275    286       BY SIMILARITY.
FT   DISULFID   280    296       BY SIMILARITY.
FT   DISULFID   298    307       BY SIMILARITY.
FT   DISULFID   373    384       BY SIMILARITY.
FT   DISULFID   378    394       BY SIMILARITY.
FT   DISULFID   396    405       BY SIMILARITY.
FT   DISULFID   411    422       BY SIMILARITY.
FT   DISULFID   416    431       BY SIMILARITY.
FT   DISULFID   433    442       BY SIMILARITY.
FT   DISULFID   450    461       BY SIMILARITY.
FT   DISULFID   455    467       BY SIMILARITY.
FT   DISULFID   469    478       BY SIMILARITY.
FT   CARBOHYD   244    244       POTENTIAL.
FT   CARBOHYD   245    245       POTENTIAL.
FT   CARBOHYD   333    333       POTENTIAL.
FT   CARBOHYD   381    381       POTENTIAL.
FT   CARBOHYD   609    609       POTENTIAL.
FT   CARBOHYD   675    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  144078 MW;  B468B73D CRC32;
```

```
  Query Match          12.5%;  Score 239;  DB 1;  Length 1295;
  Best Local Similarity  31.5%;  Pred. No. 2.03e-34;
  Matches   40;  Conservative   28;  Mismatches  50;  Indels   9;  Gaps   9;

Db    112 TPLFSGVNPCDSDPCNNGLCYPFY-GGFQCICNNGYGGSYCEE-GIDHCAQNECAEGSTC 169
          | :::| :||    | :| | |   :|| | |::|  |: |::   | |  | |  | ||
Qy     41 TGILPGCEPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TC 99

Db    170 VN-SVYNYYCDCPIGKSGRYCERTECALMG-N-I-CNHGRCIPNRDEDKNFRCVCDSGYE 225
          :  : ::| | |  | :|  |:  |  :     | |:||:|  :      : | |:||:
Qy    100 LPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLS-GVGQPY-CECNSGFT 157

Db    226 GEFCNKD 232
          |: |:::
Qy    158 GDSCDRE 164
```

```
RESULT   14
ID   DL_DROME      STANDARD;      PRT;   880 AA.
AC   P10041;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS DELTA PROTEIN PRECURSOR.
GN   DL.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   VAESSIN H., BREMER K.A., KNUST E., CAMPOS-ORTEGA J.A.;
RT   "The neurogenic gene Delta of Drosophila melanogaster is expressed in
RT   neurogenic territories and encodes a putative transmembrane protein
RT   with EGF-like repeats.";
RL   EMBO J. 6:3431-3440(1987).
RN   [2]
RP   SEQUENCE OF 422-621 FROM N.A.
RX   MEDLINE; 87218537.
RA   KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D., VAESSIN H.,
RA   CAMPOS-ORTEGA J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
RN   [3]
RP   PATTERN OF TRANSCRIPTION.
RX   MEDLINE; 91209246.
RA   HAENLIN M., KRAMATSCHEK B., CAMPOS-ORTEGA J.A.;
RT   "The pattern of transcription of the neurogenic gene Delta of
RT   Drosophila melanogaster.";
RL   DEVELOPMENT 110:905-914(1990).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL
CC       IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL
CC       CELL LINEAGES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
CC       OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
CC       THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
CC   -!- NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN LEVEL, IT IS
CC       CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY COMPETE
CC       FOR BINDING WITH THE NOTCH PROTEIN.
CC   -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: TO DROSOPHILA SERRATE PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
```

```
DR   EMBL; X06289; G7853; -.
DR   EMBL; X05140; G929563; -.
DR   PIR; S00670; S00670.
DR   PIR; A26637; A26637.
DR   FLYBASE; FBgn0000463; Dl.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PFAM; PF00008; EGF; 8.
DR   HSSP; P00740; 1IXA.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; TRANSMEMBRANE;
KW   EGF-LIKE DOMAIN; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   880       DELTA PROTEIN.
FT   DOMAIN      19   653       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   654   677       POTENTIAL.
FT   DOMAIN     678   880       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     227   258       EGF-LIKE 1.
FT   DOMAIN     256   289       EGF-LIKE 2.
FT   DOMAIN     291   329       EGF-LIKE 3.
FT   DOMAIN     331   372       EGF-LIKE 4.
FT   DOMAIN     374   416       EGF-LIKE 5.
FT   DOMAIN     418   451       EGF-LIKE 6.
FT   DOMAIN     453   489       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     491   527       EGF-LIKE 8.
FT   DOMAIN     529   565       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID   231   240       BY SIMILARITY.
FT   DISULFID   235   246       BY SIMILARITY.
FT   DISULFID   248   257       BY SIMILARITY.
FT   DISULFID   260   271       BY SIMILARITY.
FT   DISULFID   266   277       BY SIMILARITY.
FT   DISULFID   279   288       BY SIMILARITY.
FT   DISULFID   295   307       BY SIMILARITY.
FT   DISULFID   301   317       BY SIMILARITY.
FT   DISULFID   319   328       BY SIMILARITY.
FT   DISULFID   335   348       BY SIMILARITY.
FT   DISULFID   342   360       BY SIMILARITY.
FT   DISULFID   362   371       BY SIMILARITY.
FT   DISULFID   378   388       BY SIMILARITY.
FT   DISULFID   383   404       BY SIMILARITY.
FT   DISULFID   406   415       BY SIMILARITY.
FT   DISULFID   422   433       BY SIMILARITY.
FT   DISULFID   427   439       BY SIMILARITY.
FT   DISULFID   441   450       BY SIMILARITY.
FT   DISULFID   457   468       BY SIMILARITY.
FT   DISULFID   462   477       BY SIMILARITY.
FT   DISULFID   479   488       BY SIMILARITY.
FT   DISULFID   495   506       BY SIMILARITY.
FT   DISULFID   500   515       BY SIMILARITY.
FT   DISULFID   517   526       BY SIMILARITY.
FT   DISULFID   533   544       BY SIMILARITY.
FT   DISULFID   538   553       BY SIMILARITY.
FT   DISULFID   555   564       BY SIMILARITY.
FT   CARBOHYD    98    98       POTENTIAL.
FT   CARBOHYD   137   137       POTENTIAL.
FT   CARBOHYD   167   167       POTENTIAL.
FT   CARBOHYD   649   649       POTENTIAL.
FT   CONFLICT   437   438       GK -> ET (IN REF. 2).
FT   CONFLICT   459   459       G -> A (IN REF. 2).
FT   CONFLICT   490   490       S -> T (IN REF. 2).
SQ   SEQUENCE   880 AA;  94643 MW;  E967E662 CRC32;

  Query Match          12.3%;  Score 234;  DB 1;  Length 880;
  Best Local Similarity  33.3%;  Pred. No. 3.45e-33;
  Matches   53;  Conservative   30;  Mismatches  60;  Indels  16;  Gaps  12;

Db    226 CHIPKCAKGCEHGHCDKPNQCVCQLGWKGALCNECVLEP---N-CIHGTC---NK-PWTC 277
          ||  || ||:  :   | |: ||:|:||::  :|  | |:||||  |  : :|
Qy     50 CHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSC 109

Db    278 ICNEGWGGLYCNQDLN-YCTNHR-PCKNGGTCFNTGEGL-YTCKCAPGYSGDDCENEIYS 334
          | || ||: |::: : :   :   ||:|  |  :| |  | | | :|::|| |: || |

Qy    110 KCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCDREI-S 166

Db    335 CDAD-VNPC-QNGGTCIDEPHTKTGYKCHCRNGWSGKMC 371
          | :: :   |:        ||  :  || | :|  |
Qy    167 CRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQC 205


RESULT  15
ID   FBP1_STRPU     STANDARD;      PRT;  1064 AA.
AC   P10079;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FIBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE   (UEGF-1).
GN   EGF1.
OS   STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC   EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC   EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC   STRONGYLOCENTROTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90112459.
RA   DELGADILLO-REYNOSO M.G., ROLLO D.R., HURSH D.A., RAFF R.A.;
RT   "Structural analysis of the uEGF gene in the sea urchin
RT   strongylocentrotus purpuratus reveals more similarity to vertebrate
RT   than to invertebrate genes with EGF-like repeats.";
RL   J. MOL. EVOL. 29:314-327(1989).
RN   [2]
RP   SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX   MEDLINE; 87319677.
RA   HURSH D.A., ANDREWS M.E., RAFF R.A.;
RT   "A sea urchin gene encodes a polypeptide homologous to epidermal
RT   growth factor.";
RL   SCIENCE 237:1487-1490(1987).
RN   [3]
RP   AVIDIN-LIKE DOMAIN.
RX   MEDLINE; 89196806.
RA   HUNT L.T., BARKER W.C.;
RT   "Avidin-like domain in an epidermal growth factor homolog from a sea
RT   urchin.";
RL   FASEB J. 3:1760-1764(1989).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 91285254.
RA   BISGROVE B.W., ANDREWS M.E., RAFF R.A.;
RT   "Fibropellins, products of an EGF repeat-containing gene, form a
RT   unique extracellular matrix structure that surrounds the sea urchin
RT   embryo.";
RL   DEV. BIOL. 146:89-99(1991).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC       OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC       THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC       EMBRYOS AND EARLY LARVAE.
CC   -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC       DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC       LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC       MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC       AND ZYGOTICALLY.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (IA AND IB) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF REPEATS.
CC   -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08692; G161467; -.
DR   EMBL; L08692; G161466; -.
DR   EMBL; X17530; G667061; -.
DR   EMBL; M17421; G552260; -.
DR   EMBL; X17533; G667062; -.
DR   PIR; A29316; A29316.
DR   PROSITE; PS00010; ASX_HYDROXYL; 19.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 19.
DR   PFAM; PF00008; EGF; 21.
DR   PFAM; PF00431; CUB; 1.
DR   HSSP; P01132; 1EPH.
KW   BIOTIN; ALTERNATIVE SPLICING; EGF-LIKE DOMAIN; REPEAT; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1064       FIBROPELLIN I.
FT   DOMAIN      20     55       EGF-LIKE 1.
FT   DOMAIN      62    175       CUB.
FT   DOMAIN     176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     366    402       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     442    478       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     480    516       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     518    554       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     556    592       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     594    630       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632    668       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     670    706       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     708    744       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     746    782       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     784    820       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     822    858       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     860    896       EGF-LIKE 20.
FT   DOMAIN     898    934       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     936   1064       AVIDIN-LIKE.
FT   DISULFID    23     34       BY SIMILARITY.
FT   DISULFID    28     43       BY SIMILARITY.
FT   DISULFID    45     54       BY SIMILARITY.
FT   DISULFID   180    191       BY SIMILARITY.
FT   DISULFID   185    200       BY SIMILARITY.
FT   DISULFID   202    211       BY SIMILARITY.
FT   DISULFID   218    229       BY SIMILARITY.
FT   DISULFID   223    238       BY SIMILARITY.
FT   DISULFID   240    249       BY SIMILARITY.
FT   DISULFID   256    267       BY SIMILARITY.
FT   DISULFID   261    276       BY SIMILARITY.
FT   DISULFID   278    287       BY SIMILARITY.
FT   DISULFID   294    305       BY SIMILARITY.
FT   DISULFID   299    314       BY SIMILARITY.
FT   DISULFID   316    325       BY SIMILARITY.
FT   DISULFID   332    343       BY SIMILARITY.
FT   DISULFID   337    352       BY SIMILARITY.
FT   DISULFID   354    363       BY SIMILARITY.
FT   DISULFID   370    381       BY SIMILARITY.
FT   DISULFID   375    390       BY SIMILARITY.
FT   DISULFID   392    401       BY SIMILARITY.
FT   DISULFID   408    419       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   430    439       BY SIMILARITY.
FT   DISULFID   446    457       BY SIMILARITY.
FT   DISULFID   451    466       BY SIMILARITY.
FT   DISULFID   468    477       BY SIMILARITY.
FT   DISULFID   484    495       BY SIMILARITY.
FT   DISULFID   489    504       BY SIMILARITY.
FT   DISULFID   506    515       BY SIMILARITY.
FT   DISULFID   522    533       BY SIMILARITY.
FT   DISULFID   527    542       BY SIMILARITY.
FT   DISULFID   544    553       BY SIMILARITY.
FT   DISULFID   560    571       BY SIMILARITY.
FT   DISULFID   565    580       BY SIMILARITY.
FT   DISULFID   582    591       BY SIMILARITY.
FT   DISULFID   598    609       BY SIMILARITY.
FT   DISULFID   603    618       BY SIMILARITY.
FT   DISULFID   620    629       BY SIMILARITY.
FT   DISULFID   636    647       BY SIMILARITY.
FT   DISULFID   641    656       BY SIMILARITY.
FT   DISULFID   658    667       BY SIMILARITY.
FT   DISULFID   674    685       BY SIMILARITY.
FT   DISULFID   679    694       BY SIMILARITY.
FT   DISULFID   696    705       BY SIMILARITY.
FT   DISULFID   712    723       BY SIMILARITY.
FT   DISULFID   717    732       BY SIMILARITY.
FT   DISULFID   734    743       BY SIMILARITY.
FT   DISULFID   750    761       BY SIMILARITY.
FT   DISULFID   755    770       BY SIMILARITY.
FT   DISULFID   772    781       BY SIMILARITY.
FT   DISULFID   788    799       BY SIMILARITY.
FT   DISULFID   793    808       BY SIMILARITY.
FT   DISULFID   810    819       BY SIMILARITY.
FT   DISULFID   826    837       BY SIMILARITY.
FT   DISULFID   831    846       BY SIMILARITY.
FT   DISULFID   848    857       BY SIMILARITY.
FT   DISULFID   864    875       BY SIMILARITY.
FT   DISULFID   869    884       BY SIMILARITY.
FT   DISULFID   886    895       BY SIMILARITY.
FT   DISULFID   902    913       BY SIMILARITY.
FT   DISULFID   907    922       BY SIMILARITY.
FT   DISULFID   924    933       BY SIMILARITY.
FT   VARSPLIC   477    780       MISSING (IN FORM IB).
FT   CARBOHYD    30     30       POTENTIAL.
FT   CARBOHYD   136    136       POTENTIAL.
FT   CARBOHYD   851    851       POTENTIAL.
FT   CONFLICT   279    279       L -> S (IN REF. 2).
SQ   SEQUENCE   1064 AA;  112072 MW;  FBD10D48 CRC32;

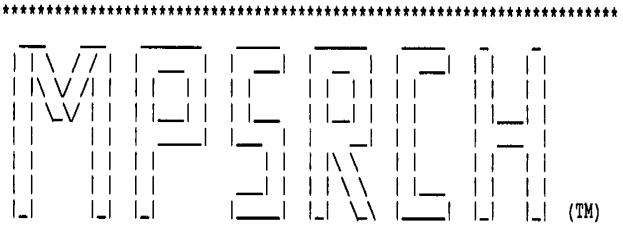
  Query Match          12.1%;  Score 230;  DB 1;  Length 1064;
  Best Local Similarity  28.3%;  Pred. No. 3.30e-32;
  Matches   36;  Conservative   29;  Mismatches  51;  Indels  11;  Gaps  10;

Db    710 DECASAPCQNGGVC-VDGVNGYVCNCAPGYTGDNCETEIDE-CASMPCLNGGACIEMVN- 766
          : |    | :|  |   : :|: |:|  |  |  |:   :: | :  |::| :|: :
Qy     48 EPCHKKVCAHG-CCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAF 105

Db    767 GYTCQCVAGYTGVIC--ETDI-DECASAPCQNGGVCTDTIN-GYICACVPGFTGSNCETN 822
          :|:| |: |  ||:|  | |: : |    | :|     :   | | | :|:|| :|: :
Qy    106 SYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDRE 164

Db    823 IDECASD 829
          |  | ::
Qy    165 IS-CRGE 170


Search completed: Fri May 28 09:41:43 1999
Job time : 21 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 28 09:42:02 1999; MasPar time 17.34 Seconds
764.994 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-14
Description: (1-243) from US09191647.pep
Perfect Score: 1905
Sequence: 1 ILDVASLRQAPGENGTSFHG.....SSFVDEVEKVVKCGCARCAS 243

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 40.603; Variance 58.453; scale 0.695

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1060	55.6	1531	11	O88279	MEGF4.	1.45e-265
2	990	52.0	739	4	O75094	MEGF5 (FRAGMENT).	2.96e-245
3	988	51.9	1523	11	O88280	MEGF5.	1.12e-244
4	405	21.3	79	4	O75093	MEGF4 (FRAGMENT).	2.61e-79
5	294	15.4	2447	13	O13149	NOTCH 2 (FRAGMENT).	1.37e-49
6	290	15.2	1203	11	Q06008	NOTCH PROTEIN HOMOLOG	1.52e-48
7	290	15.2	2470	11	O35516	CELL SURFACE PROTEIN.	1.52e-48
8	270	14.2	1722	5	Q19350	SIMILAR TO EGF-LIKE RE	2.38e-43
9	269	14.1	1219	11	Q63722	JAGGED PROTEIN.	4.30e-43
10	260	13.6	752	13	O42374	NOTCH RECEPTOR PROTEIN	8.90e-41
11	258	13.5	1218	4	O14902	TRANSMEMBRANE PROTEIN	2.90e-40
12	258	13.5	1218	4	Q15816	TRANSMEMBRANE PROTEIN	2.90e-40
13	258	13.5	1218	4	O15122	JAGGED1.	2.90e-40
14	258	13.5	1227	4	P78504	JAGGED 1 (TRANSMEMBRAN	2.90e-40
15	256	13.4	1193	13	Q90819	C-SERATE-1 PROTEIN (FR	9.42e-40
16	256	13.4	1212	13	O42347	C-SERRATE-2 (FRAGMENT)	9.42e-40
17	254	13.3	530	5	Q24526	SLIT LOCUS ENCODING A	3.06e-39
18	252	13.2	2653	5	Q25253	NOTCH HOMOLOG SCALLOPE	9.91e-39
19	250	13.1	1687	11	Q61204	NOTCH2-LIKE (EGF REPEA	3.21e-38
20	244	12.8	2352	5	O61240	HRNOTCH PROTEIN.	1.07e-36
21	243	12.8	2531	5	O16004	NOTCH HOMOLOG.	1.92e-36
22	240	12.6	406	5	Q25059	FIBROPELLIN III (FRAGM	1.11e-35
23	240	12.6	1964	11	O35442	NOTCH4.	1.11e-35
24	237	12.4	601	5	Q20204	F40E10.4 PROTEIN (FRAG	6.33e-35
25	236	12.4	1095	4	Q99458	NOTCH4 (FRAGMENT).	1.13e-34
26	236	12.4	1999	4	Q99940	NOTCH4.	1.13e-34
27	236	12.4	2003	4	O00306	NOTCH4.	1.13e-34
28	234	12.3	473	5	Q25464	ADHESIVE PLAQUE MATRIX	3.60e-34
29	233	12.2	832	5	Q99108	NEUROGENIC LOCUS DELTA	6.43e-34
30	233	12.2	4590	4	Q14517	CADHERIN-RELATED TUMOR	6.43e-34
31	231	12.1	434	11	O55139	JAGGED2 PROTEIN (FRAGM	2.04e-33
32	231	12.1	518	11	O70219	JAGGED 2 (JAGGED 2 PRO	2.04e-33
33	230	12.1	529	5	Q25058	FIBROPELLIN IA (FRAGME	3.64e-33
34	230	12.1	661	5	O61537	SPERM TRANSMEMBRANE PR	3.64e-33
35	228	12.0	955	4	Q99466	NOTCH4 (FRAGMENT).	1.15e-32
36	226	11.9	1202	11	P97607	JAGGED2 (FRAGMENT).	3.65e-32
37	227	11.9	1476	13	Q90285	PUTATIVE EXTRACELLULAR	2.05e-32
38	223	11.7	762	13	O42373	NOTCH RECEPTOR PROTEIN	2.04e-31
39	219	11.5	721	13	Q91902	X-DELTA-1.	2.02e-30
40	219	11.5	3871	5	Q20911	ZC116.3 PROTEIN.	2.02e-30
41	218	11.4	717	13	P87357	DELTAD TRANSMEMBRANE P	3.57e-30
42	216	11.3	642	13	P79941	NOTCH LIGAND X-DELTA-2	1.12e-29
43	216	11.3	728	13	Q90656	TRANSMEMBRANE PROTEIN	1.12e-29
44	216	11.3	802	13	O57462	DELTAA.	1.12e-29
45	214	11.2	263	4	Q99740	SOLUBLE PROTEIN JAGGED	3.49e-29

ALIGNMENTS

RESULT 1
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGF4.
GN MEGF4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL GENOMICS 51:27-34(1998).
DR EMBL; AB011530; D1033423; -.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 1531 AA; 167497 MW; 5C5EBDF4 CRC32;

Query Match 55.6%; Score 1060; DB 11; Length 1531;
Best Local Similarity 54.6%; Pred. No. 1.45e-265;
Matches 125; Conservative 45; Mismatches 59; Indels 0; Gaps 0;

```
Db  1303 NGTSFHGCIRNLYINNELQDFTKTQMKPGVVPGCEPCRKLYCLHGICQPNATPGPVCHCE 1362
         ||||||||||||||:|||||  |   |  |::||||||:|  | || |||:: :|  | ||
Qy     14 NGTSFHGCIRNLYINSELQDFRKMPMQTGILPGCEPCHKKVCAHGCCQPSSQSGFTCECE 73

Db  1363 AGWGGLHCDQPVDGPCHGHKCVHGKCVPLDALAYSCQCQDGYSGALCNQVGAVAEPCGGL 1422
           || |  |||  : || |:|||||| |:|::|::|||| | :| :| |||::   : :||  :
Qy     74 EGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQMI 133

Db  1423 QCLHGHCQASATRGAHCVCSPGFSGELCEQESECRGDPVRDFHRVQRGYAICQTTRPLSW 1482
           | || |: |:   : | |::||:|: |::|  |||: :||: : |:||| ||||: :|:
Qy    134 KCKHGKCRLSGVGQPYCECNSGFTGDSCDREISCRGERIRDYYQKQQGYAACQTTKKVSR 193

Db  1483 VECRGACPGQGCCQGLRLKRRKLTFECSDGTSFAEEVEKPTKCGCAPCA 1531
          :||||:|:|  ||  || |||| :|||:||:|| :||||  ||||| ||
```

```
Qy   194 LECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCARCA 242


RESULT   2
ID   O75094      PRELIMINARY;      PRT;   739 AA.
AC   O75094;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5 (FRAGMENT).
GN   MEGF5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011538; D1033429; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   739 AA;  80364 MW;  DC6BCB63 CRC32;

  Query Match           52.0%;  Score 990;  DB 4;  Length 739;
  Best Local Similarity  51.3%;  Pred. No. 2.96e-245;
  Matches  123;  Conservative   50;  Mismatches  62;  Indels   5;  Gaps   4;

Db   504 LSALRQGTDRPLGGFHGCIHEVRINNELQDFKALPPQSLGVSPGCKSC--TVCKHGLCRS 561
         :::|||:      :|||||::: ||:|||||: :| |: |: ||| :|   || || |::
Qy     4 VASLRQAPGENGTSFHGCIRNLYINSELQDFRKMPMQT-GILPGCEPCHKKVCAHGCCQP 62

Db   562 VEKDSVVCECRPGWTGPLCDQEARDPCLGHRCHHGKCVATGT-SYMCKCAEGYGGDLCDN 620
             :  |||  || |||||| : |||||::| || |::  : || ||| || || |||:
Qy    63 SSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDE 122

Db   621 KNDSANACSAFKCHHGQCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGY 680
           :|  |:|  :|| || |::|  |:||| | :||:|: |::| :| |: :|:  ::| ||
Qy   123 EEDLFNPCQMIKCKHGKCRLSGVGQPYCECNSGFTGDSCDREISCRGERIRDYYQKQQGY 182

Db   681 ASCATASKVPIMECRGGC-GPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS 739
         |:| |: ||: :|||||| | ||| | ||||||| |:||||||||:|||: : |||  |:
Qy   183 AACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCARCA 242


ULT   3
ID   O88280      PRELIMINARY;      PRT;  1523 AA.
AC   O88280;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF5.
GN   MEGF5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011531; D1033424; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1523 AA;  167767 MW;  2BD845D0 CRC32;

  Query Match           51.9%;  Score 988;  DB 11;  Length 1523;
  Best Local Similarity  50.8%;  Pred. No. 1.12e-244;
  Matches  122;  Conservative   50;  Mismatches  63;  Indels   5;  Gaps   4;

Db  1288 LSALRQGADRPLGGFHGCIHEVRINNELQDFKALPPQSLGVSPGCKSC--TVCRHGLCRS 1345
         :::|||::     :|||||::: ||:|||||: :| |: |: ||| :|   || || |::
Qy     4 VASLRQAPGENGTSFHGCIRNLYINSELQDFRKMPMQT-GILPGCEPCHKKVCAHGCCQP 62

Db  1346 VEKDSVVCECHPGWTGPLCDQEAQDPCLGHSCSHGTCVATGN-SYVCKCAEGYEGPLCDQ 1404
               :  |||  || |||||| : |||||: | ||||::    || ||| ||  | |||:
Qy    63 SSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDE 122

Db  1405 KNDSANACSAFKCHHGQCHISDRGEPYCLCQPGFSGNHCEQENPCLGEIVREAIRRQKDY 1464
           :|  |:|  :|| || |::|  |:||| | :||:|: |::| :| || :|:  ::|  |
Qy   123 EEDLFNPCQMIKCKHGKCRLSGVGQPYCECNSGFTGDSCDREISCRGERIRDYYQKQQGY 182

Db  1465 ASCATASKVPIMVCRGGC-GSQCCQPIRSKRRKYVFQCTDGSSFVEEVERHLECGCRECS 1523
         |:| |: ||: : ||||| |:||| |:||||||| |:||||||||:|||: : |||  |:
Qy   183 AACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCARCA 242


RESULT   4
ID   O75093      PRELIMINARY;      PRT;    79 AA.
AC   O75093;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF4 (FRAGMENT).
GN   MEGF4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011537; D1033428; -.
DR   PROSITE; PS01185; CTCK_1; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   79 AA;  8809 MW;  96C95FFE CRC32;

  Query Match           21.3%;  Score 405;  DB 4;  Length 79;
  Best Local Similarity  60.8%;  Pred. No. 2.61e-79;
  Matches   48;  Conservative   15;  Mismatches  16;  Indels   0;  Gaps   0;

Db     1 ESECRGDPVRDFHQVQRGYAICQTTRPLSWVECRGSCPGQGCCQGLRLKRRKFTFECSDG 60
           |  |||: :||: | |:||| |||| : :|::||||:|:|  ||  || ||||::|||:||
Qy   164 EISCRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDG 223

Db    61 TSFAEEVEKPTKCGCALCA 79
           :|| :|||| ||||| ||
Qy   224 SSFVDEVEKVVKCGCARCA 242


RESULT   5
ID   O13149      PRELIMINARY;      PRT;  2447 AA.
AC   O13149;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH 2 (FRAGMENT).
OS   FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   TETRAODONTIFORMES; TETRAODONTOIDEI; TETRAODONTIDAE; FUGU.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NAKAMURA T., TROWSDALE J.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB004829; D1021371; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   2447 AA;  262542 MW;  3CDA4F7A CRC32;

  Query Match           15.4%;  Score 294;  DB 13;  Length 2447;
  Best Local Similarity  41.9%;  Pred. No. 1.37e-49;
  Matches   49;  Conservative   22;  Mismatches  35;  Indels  11;  Gaps  11;

Db    381 CEHGGQC-VNTEGSFTCNCAKGYAGPRCEQDVNE-CASNPCQNDGTCLDRIG-DYSCICM 437
            | ||  |  :::::|||:|  |  || |:|  |: | :| | : ||||   :  ||| |:
Qy     55 CAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFSYSCKCL 112

Db    438 PGFGGTHC-ENE-L-NECLSSPCLNRGKC-LDQVSRFVCECPAGFSGEMCQIDIDEC 490
            | ||  | |:| | | |    | ::||| |  |::  ||| :||:|: |: :|  |
Qy    113 EGHGGVLCDEEEDLFNPCQMIKC-KHGKCRLSGVGQPYCECNSGFTGDSCDREIS-C 167


RESULT    6
ID   Q06008      PRELIMINARY;      PRT;  1203 AA.
AC   Q06008;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH PROTEIN HOMOLOG 2 (MOTCH B PROTEIN) (FRAGMENT).
GN   NOTCH2 OR MOTCH B.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
DR   EMBL; X68279; G287990; -.
DR   MGD; MGI:97364; NOTCH2.
DR   PFAM; PF00008; EGF; 27.
DR   PFAM; PF00066; notch; 1.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT.
FT   NON_TER       1      1
FT   NON_TER    1203   1203
SQ   SEQUENCE   1203 AA;  128982 MW;  A5A95551 CRC32;

  Query Match           15.2%;  Score 290;  DB 11;  Length 1203;
  Best Local Similarity  33.3%;  Pred. No. 1.52e-48;
  Matches   43;  Conservative   34;  Mismatches  41;  Indels  11;  Gaps  10;

Db    255 DNCDPDPCHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 311
            : |    | || || : :| :|| |: | ||::| :: :: | :: |:: | |: :   :
Qy     48 EPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 106

Db    312 YQCNCQPGTSGLNC--EIN-FDDCASNPCMHGVC-VDGINR-YSCVCSPGFTGQRCNIDI 366
            | |:|  | :|: |  | : |: |    |:|| | : |: : | | |::|||: |: :|
Qy    107 YSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDREI 165

Db    367 DECASNPCR 375
            | ::  |
Qy    166 S-CRGERIR 173


RESULT    7
ID   O35516      PRELIMINARY;      PRT;  2470 AA.
AC   O35516;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CELL SURFACE PROTEIN.
GN   NOTCH2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   EXP. CELL RES. 204:364-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B/6; TISSUE=THYMUS;
RA   HAMADA Y., HIGUCHI M., TSUJIMOTO Y.;
RL   SUBMITTED (JUL-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D32210; D1022953; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   2470 AA;  265325 MW;  CA94E03A CRC32;

  Query Match           15.2%;  Score 290;  DB 11;  Length 2470;
  Best Local Similarity  33.3%;  Pred. No. 1.52e-48;
  Matches   43;  Conservative   34;  Mismatches  41;  Indels  11;  Gaps  10;

Db    570 DNCDPDPCHHGQCQDGIDS-YTCICNPGYMGAICSDQIDE-CYSSPCLNDGRCIDLVN-G 626
            : |    | || || : :| :|| |: | ||::| :: :: | :: |:: | |: :   :
Qy     48 EPCHKKVCAHGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLPINAFS 106

Db    627 YQCNCQPGTSGLNC--EIN-FDDCASNPCMHGVC-VDGINR-YSCVCSPGFTGQRCNIDI 681
            | |:|  | :|: |  | : |: |    |:|| | : |: : | | |::|||: |: :|
Qy    107 YSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPY-CECNSGFTGDSCDREI 165

Db    682 DECASNPCR 690
            | ::  |
Qy    166 S-CRGERIR 173


RESULT    8
ID   Q19350      PRELIMINARY;      PRT;  1722 AA.
AC   Q19350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO EGF-LIKE REPEATS. NCBI GI: 1125776.
GN   F11C7.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
```

```
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A., VETTER J.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U42839; G1125776; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 19.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PFAM; PF00008; EGF; 24.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1722 AA;  188383 MW;  CCFB86B8 CRC32;

  Query Match          14.2%;  Score 270;  DB 5;  Length 1722;
  Best Local Similarity  29.1%;  Pred. No. 2.38e-43;
  Matches   37;  Conservative   34;  Mismatches  48;  Indels   8;  Gaps   8;

Db   1252 DECEGVECHNGGKCVKNRSEKIVCQCGNSWMGDSCNVTKTTNCKDSPCQNFGQCMQKTDT 1311
           : |    | :|  |  :    : |:| ::|||  |:      |  : | : | |:   ::
Qy     48 EPCHKKVCAHGC-CQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVH-GTCLP-INA 104

Db   1312 F-FECNCMDGYSGELC-EQRDV-NECNHYDCNRGHCVMT-VSGPACQCEMGYTGRFCEKL 1367
           | : |:|::| :| || |: |: | |    |::| | :: |: | |:|: |:||  |::
Qy    105 FSYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHGKCRLSGVGQPYCECNSGFTGDSCDRE 164

Db   1368 LNQCSSN 1374
           :: | ::
Qy    165 IS-CRGE 170


RESULT    9
ID   Q63722      PRELIMINARY;      PRT;  1219 AA.
AC   Q63722; P70640;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   JAGGED PROTEIN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCIATIC NERVE;
RX   MEDLINE; 95211842.
RA   LINDSELL C.E., SHAWBER C.J., BOULTER J., WEINMASTER G.;
RT   "Jagged: a mammalian ligand that activates Notch1.";
RL   CELL 80:909-917(1995).
DR   EMBL; L38483; G1492111; -.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1219 AA;  134325 MW;  B193F948 CRC32;

  Query Match          14.1%;  Score 269;  DB 11;  Length 1219;
  Best Local Similarity  37.8%;  Pred. No. 4.30e-43;
  Matches   45;  Conservative   19;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 NVCGPHGKCKSESGGKFTCDCNKGFTGTYCHENIND-CEGNPCTNGGTCID-GVNSYKCI 654
           :||: || | : | : |||:|: |  |  | :  || | || | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 CSDGWEGAHC--ENNI-NDCSQNPCHYGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 709
           -| :|  |  |  |::: | |    |  |  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT   10
ID   O42374      PRELIMINARY;      PRT;  752 AA.
AC   O42374;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOTCH RECEPTOR PROTEIN (FRAGMENT).
GN   NOTCH6.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WESTIN J., LARDELLI M.;
RL   DEV. GENES EVOL. 207:51-63(1997).
DR   EMBL; Y10354; E293438; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 9.
DR   PROSITE; PS01186; EGF_2; 15.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PFAM; PF00008; EGF; 16.
DR   PFAM; PF00066; notch; 2.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
FT   NON_TER     752    752
SQ   SEQUENCE   752 AA;  82103 MW;  72E254FB CRC32;

  Query Match          13.6%;  Score 260;  DB 13;  Length 752;
  Best Local Similarity  35.2%;  Pred. No. 8.90e-41;
  Matches   44;  Conservative   25;  Mismatches  43;  Indels  13;  Gaps  12;

Db     64 PCASQPCQRRGVCQPSLDYTSYTCKCHSGWEGAQCTEDKDE-CKKSPCQNGARCVNIVG- 121
           ||    | : | |||| : :::|| |  || |: | :  :: |  : | :|: |: | :
Qy     49 PCHKKVCAH-GCCQPSSQ-SGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGT-CLPINAF 105

Db    122 SYRCECPPGYSGDNCQTHID--D-CSSNPCRNGGTC-VDKVGR-YLCECRAGFYGERCEE 176
           || | |  | :|  |:   |  : |    |::|  | :  ||: | ||| :|| |: |:
Qy    106 SYSCKCLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPY-CECNSGFTGDSCDR 163

Db    177 EVDEC 181
           |: |
Qy    164 EIS-C 167


RESULT   11
ID   O14902      PRELIMINARY;      PRT;  1218 AA.
AC   O14902;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE PROTEIN JAGGED 1.
GN   HJ1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95211842.
RA   LINDSELL C.E., SHAWBER C.J., BOULTER J., WEINMASTER G.;
RT   "Jagged: a mammalian ligand that activates Notch1.";
RL   CELL 80:909-917(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   BASH J., ZONG W.-X., GELINAS C.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF028593; G2599082; -.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1218 AA;  133797 MW;  07B97EE3 CRC32;

  Query Match          13.5%;  Score 258;  DB 4;  Length 1218;
```

```
  Best Local Similarity  37.0%;  Pred. No. 2.90e-40;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 NVCGPHGKCKSQSGGKFTCDCNKGFTGTYCHENIND-CESNPCRNGGTCID-GVNSYKCI 654
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:     || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 CSDGWEGAYC--ETNI-NDCSQNPCHNGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT  12
ID   Q15816      PRELIMINARY;      PRT;  1218 AA.
AC   Q15816;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE PROTEIN JAGGED 1.
GN   HJ1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95211842.
RA   LINDSELL C.E., SHAWBER C.J., BOULTER J., WEINMASTER G.;
RT   "Jagged: a mammalian ligand that activates Notch1.";
RL   CELL 80:909-917(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   GRAY G.E., MANN R.S., MITSIADIS E., HENRIQUE D., CARACANGIU M.,
RA   ISH-HOROWICZ D., ARTAVANIS-TSAKONAS S.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U61276; G1438937; -.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1218 AA;  133738 MW;  E8D64FED CRC32;

  Query Match          13.5%;  Score 258;  DB 4;  Length 1218;
  Best Local Similarity  37.0%;  Pred. No. 2.90e-40;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 NVCGPHGKCKSQSGGKFTCDCNKGFTGTYCHENIND-CESNPCRNGGTCID-GVNSYKCI 654
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:     || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 CSDGWEGAYC--ETNI-NDCSQNPCHNGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT  13
ID   O15122      PRELIMINARY;      PRT;  1218 AA.
AC   O15122;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   JAGGED1.
GN   JAG1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ODA T., ELKAHLOUN A.G., PIKE B.L., OKAJIMA K., KRANTZ I.D., GENIN A.,
RA   PICCOLI D.A., MELTZER P.S., SPINNER N.B., COLLINS F.S.,
RA   CHANDRASEKHARAPPA S.C.;
RL   NAT. GENET. 0:0-0(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97422615.
RA   ODA T., ELKAHLOUN A.G., MELTZER P.S., CHANDRASEKHARAPPA S.C.;
RT   "Identification and cloning of the human homolog (JAG1) of the rat
RT   Jagged1 gene from the Alagille syndrome critical region at 20p12.";
RL   GENOMICS 43:376-379(1997).
DR   EMBL; AF003837; G2228793; -.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1218 AA;  133858 MW;  20F471DB CRC32;

  Query Match          13.5%;  Score 258;  DB 4;  Length 1218;
  Best Local Similarity  37.0%;  Pred. No. 2.90e-40;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 NVCGPHGKCKSQSGGKFTCDCNKGFTGTYCHENIND-CESNPCRNGGTCID-GVNSYKCI 654
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:     || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 CSDGWEGAYC--ETNI-NDCSQNPCHNGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


RESULT  14
ID   P78504      PRELIMINARY;      PRT;  1227 AA.
AC   P78504;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   JAGGED 1 (TRANSMEMBRANE PROTEIN JAGGED).
GN   HJ1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95211842.
RA   LINDSELL C.E., SHAWBER C.J., BOULTER J., WEINMASTER G.;
RT   "Jagged: a mammalian ligand that activates Notch1.";
RL   CELL 80:909-917(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LI L., DENG Y., BANTA A.B., HOOD L.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE OF 14-1227 FROM N.A.
RX   MEDLINE; 97115768.
RA   ZIMRIN A.B., PEPPER M.S., MCMAHON G., NGUYEN F., MONTESANO R.,
RA   MACIAG T.;
RT   "An antisense oligonucleotide to the notch ligand jagged enhances
RT   fibroblast growth factor-induced angiogenesis in vitro.";
RL   J. BIOL. CHEM. 271:32499-32502(1996).
RN   [4]
RP   REVISIONS TO 14-1227.
RA   ZIMRIN A.B., NGUYEN F., MACIAG T.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U73936; G1695274; -.
DR   EMBL; U77720; G2130537; -.
DR   PFAM; PF00008; EGF; 14.
KW   TRANSMEMBRANE.
FT   CONFLICT   1187   1227
FT                              QRHADKTPKLDKQTGQQRLGKCPELKPNGVHRIADRGHCR
FT                              R -> NGTPTKHPNWTNKQDNRDLESAQSLNRMEYIV
FT                              (IN REF. 1 AND 2).
SQ   SEQUENCE   1227 AA;  134770 MW;  5D300B81 CRC32;

  Query Match          13.5%;  Score 258;  DB 4;  Length 1227;
  Best Local Similarity  37.0%;  Pred. No. 2.90e-40;
  Matches   44;  Conservative   20;  Mismatches  45;  Indels  10;  Gaps   9;

Db    597 NVCGPHGKCKSQSGGKFTCDCNKGFTGTYCHENIND-CESNPCRNGGTCID-GVNSYKCI 654
```

```
          :||: || | : | : |||:|: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    655 CSDGWEGAYC--ETNI-NDCSQNPCHNGGTCR-DLVNDFYCDCKNGWKGKTCHSRDSQC 709
          | :|  |  |  | :: | |    | :|  ||   | : ||:|::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167

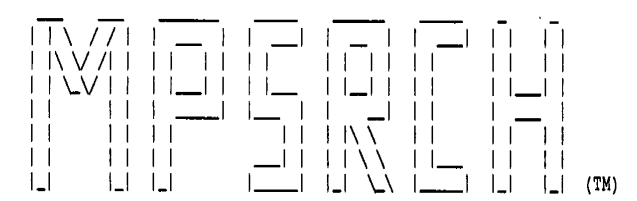

RESULT  15
ID   Q90819      PRELIMINARY;      PRT;  1193 AA.
AC   Q90819;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C-SERATE-1 PROTEIN (FRAGMENT).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 96175595.
RA   MYAT A., HENRIQUE D., ISH-HOROWICZ D., LEWIS J.;
RT   "A chick homologue of Serrate and its relationship with Notch and
RT   Delta homologues during central neurogenesis.";
RL   DEV. BIOL. 174:233-247(1996).
DR   EMBL; X95283; E224084; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 10.
DR   PROSITE; PS01186; EGF_2; 12.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER       1      1
SQ   SEQUENCE   1193 AA;  131039 MW;  55E5FCD1 CRC32;

  Query Match          13.4%;  Score 256;  DB 13;  Length 1193;
  Best Local Similarity  37.0%;  Pred. No. 9.42e-40;
  Matches   44;  Conservative   21;  Mismatches  44;  Indels  10;  Gaps   9;

Db    571 NVCGPHGKCKSQAGGKFTCECNKGFTGTYCHENIND-CESNPCKNGGTCID-GVNSYKCI 628
          :||: || | : : : |||||: |  |  | :  || | :| | :| ||:    || |
Qy     53 KVCA-HGCCQPSSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHG-TCLPINAFSYSCK 110

Db    629 CSDGWEGTYC--ETNI-NDCSKNPCHNGGTCR-DLVNDFFCECKNGWKGKTCHSRDSQC 683
          | :|  |  |  | :: | | :  | :|  ||   | : :|||::|  | :|  |:  |
Qy    111 CLEGHGGVLCDEEEDLFNPCQMIKCKHG-KCRLSGVGQPYCECNSGFTGDSCD-REISC 167


Search completed: Fri May 28 09:42:55 1999
Job time : 53 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]rch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: , Sat May 29 18:13:20 1999; MasPar time 8535.17 Seconds
1545.223 Million cell updates/sec

Tabular output not generated.

Title: >US-09-191-647-1
Description: (1-4758) from US09191647.seq
Perfect Score: 4758
N.A. Sequence: 1 atgcgcggcgttggctggca.....aaaaaaaaaaaaaaactcga 4758
Comp: tacgcgccgcaaccgaccgt.....ttttttttttttttttgagct

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 12.589; Variance 6.522; scale 1.930

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	4588	96.4	5233	31	AF055585	Homo sapiens neurogeni	0.00e+00
	2	4502	94.6	4950	29	AB017168	Homo sapiens mRNA for	0.00e+00
	3	2498	52.5	5121	32	AF074960	Mus musculus neurogeni	0.00e+00
	4	1249	26.3	5210	32	AB011531	Rattus norvegicus mRNA	0.00e+00
	5	1231	25.9	5094	29	AB017167	Homo sapiens mRNA for	0.00e+00
	6	1202	25.3	5015	29	AB017169	Homo sapiens mRNA for	0.00e+00
	7	1177	24.7	4950	32	AB011530	Rattus norvegicus mRNA	0.00e+00
	8	505	10.6	2553	31	AF075240	Homo sapiens SLIT1 pro	0.00e+00
	9	469	9.9	6921	29	AB011538	Homo sapiens mRNA for	0.00e+00
	10	182	3.8	591	32	AB017170	Rattus norvegicus mRNA	1.15e-110
	11	106	2.2	5401	21	DMSLIT	Drosophila mRNA for sl	2.11e-52
c	12	80	1.7	7218	25	I66494	Sequence 14 from paten	1.95e-33
	13	68	1.4	297	34	G41330	Z1154 Zebrafish AB Dan	5.24e-25
	14	67	1.4	721	32	AF088902	Mus musculus SLIT1 pro	2.56e-24
	15	57	1.2	7218	25	I66494	Sequence 14 from paten	1.49e-17
	16	58	1.2	98695	22	AC005556	Drosophila melanogaste	3.23e-18
	17	50	1.1	491	31	HUMZE12G05	Homo sapiens full leng	5.38e-13
	18	50	1.1	8064	32	MMNOTCHA	M.musculus notch-1 mRN	5.38e-13
	19	54	1.1	8091	30	HSU97669	Homo sapiens Notch3 (N	1.40e-15
	20	46	1.0	7943	32	MMNOTC	M.musculus mRNA for No	1.79e-10
	21	48	1.0	8221	32	RRNOTCH	R.rattus mRNA homologu	1.00e-11
	22	47	1.0	8503	22	LCU58977	Lucilia cuprina Notch	4.25e-11
c	23	49	1.0	74371	31	AC005369	Homo sapiens chromosom	2.33e-12
c	24	42	0.9	965	25	AR024229	Sequence 22 from paten	5.00e-08
	25	45	0.9	1196	30	HSNTCH03	Homo sapiens Notch3 (N	7.44e-10
	26	44	0.9	1674	21	DROSLITEGF	Drosophila melanogaste	3.06e-09
	27	41	0.9	7332	29	HUMTAN1	Human TAN-1 mRNA (homo	1.98e-07
	28	41	0.9	7471	24	BRNOTCH	B.rerio Notch mRNA.	1.98e-07
	29	41	0.9	9166	24	XELXOTCH	X.laevis Xotch protein	1.98e-07
	30	44	0.9	30088	21	CEC26G2	Caenorhabditis elegans	3.06e-09
c	31	45	0.9	41150	30	AC004663	Homo sapiens chromosom	7.44e-10
	32	37	0.8	215	25	I28278	Sequence 5 from patent	4.28e-05
	33	37	0.8	406	23	AF020290	Oryctolagus cuniculus	4.28e-05
	34	36	0.8	965	25	AR024229	Sequence 22 from paten	1.58e-04
	35	36	0.8	1056	23	MVU87256	Mustela vison GT dinuc	1.58e-04
c	36	40	0.8	1056	23	MVU87256	Mustela vison GT dinuc	7.77e-07
	37	38	0.8	2256	24	DRNOTCH6	D.rerio notch6 mRNA.	1.14e-05
	38	38	0.8	3314	29	AB011537	Homo sapiens mRNA for	1.14e-05
	39	39	0.8	8287	32	RATNOTCHX	Rat notch 2 mRNA.	3.00e-06
	40	38	0.8	13692	21	CELLIN12A	C.elegans homeotic lin	1.14e-05
c	41	38	0.8	40970	21	CER107	Caenorhabditis elegans	1.14e-05
	42	35	0.7	2087	29	HUMPGH3A	Human pgH3 mRNA for pr	5.71e-04
	43	35	0.7	3162	30	AF003522	Homo sapiens Delta mRN	5.71e-04
c	44	35	0.7	10772	21	AF012089	Drosophila melanogaste	5.71e-04
	45	35	0.7	17137	21	DRONOTCH03	D.melanogaster Notch 1	5.71e-04

ALIGNMENTS

RESULT 1
LOCUS AF055585 5233 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens neurogenic extracellular slit protein Slit2 mRNA, complete cds.
ACCESSION AF055585
NID g4151204
VERSION AF055585.1 GI:4151204
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5233)
AUTHORS Holmes,G.P., Negus,K., Burridge,L., Raman,S., Algar,E., Yamada,T. and Little,M.H.
TITLE Distinct but overlapping expression patterns of two vertebrate slit homologs implies functional roles in CNS development and organogenesis
JOURNAL Mech. Dev. 79 (1-2), 57-72 (1998)
REFERENCE 2 (bases 1 to 5233)
AUTHORS Holmes,G.P., Burridge,L., Negus,K., Raman,S. and Little,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1998) Center for Molecular and Cellular Biology, University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia
FEATURES Location/Qualifiers
source 1. .5233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4p16.3"
/dev_stage="fetal"
/tissue_type="kidney; brain"
CDS 557. .5122
/note="leucine-rich repeat EGF-like repeat secreted protein; similar to Drosophila neurogenic extracellular slit protein; expressed in developing CNS, limbs and metanephros"

/codon_start=1
/product="neurogenic extracellular slit protein Slit2"
/protein_id="AAD04309.1"
/db_xref="PID:g4151205"
/db_xref="GI:4151205"
/translation="MRGVGWQMLSLSLGLVLAILNKVAPQACPAQCSCSGSTVDCHGL
ALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDL
KELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIPRKAFRGAVDIKNLQLDY
NQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRTFRLHSNNLYCDCHLA
WLSDWLRQRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSGHQSFMAPSCSVLHCPAA
CTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNN
QISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLLNANKINCLRVDAF
QDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLKWLADYLHTNPI
ETSGARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACPEKCRCEGTT
VDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFSNNKITDIE
EGAFEGASGVNEILLTSNRLENVQHKMFKGLEKPQNLMLRSNRITCVGNDSFIGLSSV
RMLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKKRIVTGNPR
CQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRCSNKGLKVL
PKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQSFSNMTQLLTL
ILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLYCD
CNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPC
LSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTCHLKEGEED
GFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD
FCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVN
GYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQCLPGYQGE
KCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYR
GRVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQ
STLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDFQKVPMQTG
ILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCL
PINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYCECSSGYTG
DSCDREISCRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKY
SFECTDGSSFVDEVEKVVKCGCTRCVS"
misc_feature 638. .730
/note="Region: amino-flanking"
misc_feature 734. .766
/note="Region: leucine-rich region 1-1"
misc_feature 779. .838
/note="Region: leucine-rich region 1-2"
misc_feature 851. .907
/note="Region: leucine-rich region 1-3"
misc_feature 923. .982
/note="Region: leucine-rich region 1-4"
misc_feature 995. .1054
/note="Region: leucine-rich region 1-5"
misc_feature 1067. .1126
/note="Region: leucine-rich region 1-6"
misc_feature 1139. .1174
/note="Region: leucine-rich region 1-7"
misc_feature 1181. .1372
/note="Region: carboxy-flanking"
misc_feature 1373. .1465
/note="Region: amino-flanking"
misc_feature 1469. .1501
/note="Region: leucine-rich region 2-1"
misc_feature 1514. .1573
/note="Region: leucine-rich region 2-2"
misc_feature 1586. .1645
/note="Region: leucine-rich region 2-3"
misc_feature 1658. .1717
/note="Region: leucine-rich region 2-4"
misc_feature 1730. .1789
/note="Region: leucine-rich region 2-5"
misc_feature 1802. .1837
/note="Region: leucine-rich region 2-6"
misc_feature 1844. .2047
/note="Region: carboxy-flanking"
misc_feature 2048. .2140
/note="Region: amino-flanking"
misc_feature 2144. .2176
/note="Region: leucine-rich region 3-1"
misc_feature 2192. .2251
/note="Region: leucine-rich region 3-1A"
misc_feature 2264. .2323
/note="Region: leucine-rich region 3-2"
misc_feature 2336. .2395
/note="Region: leucine-rich region 3-3"
misc_feature 2408. .2467
/note="Region: leucine-rich region 3-4"
misc_feature 2480. .2515
/note="Region: leucine-rich region 3-5"
misc_feature 2522. .2710
/note="Region: carboxy-flanking"
misc_feature 2711. .2803
/note="Region: amino-flanking"
misc_feature 2807. .2839
/note="Region: leucine-rich region 4-1"
misc_feature 2849. .2908
/note="Region: leucine-rich region 4-2"
misc_feature 2921. .2980
/note="Region: leucine-rich region 4-3"
misc_feature 2993. .3052
/note="Region: leucine-rich region 4-4"
misc_feature 3065. .3100
/note="Region: leucine-rich region 4-5"
misc_feature 3107. .3280
/note="Region: carboxy-flanking"
misc_feature 3284. .3397
/note="Region: EGF-like repeat 1"
misc_feature 3401. .3520
/note="Region: EGF-like repeat 2"
misc_feature 3524. .3634
/note="Region: EGF-like repeat 3"
misc_feature 3638. .3754
/note="Region: EGF-like repeat 4"
misc_feature 3758. .3868
/note="Region: EGF-like repeat 5"
misc_feature 3893. .4003
/note="Region: EGF-like repeat 6"
misc_feature 4520. .4636
/note="Region: EGF-like repeat 6A"
misc_feature 4643. .4753
/note="Region: EGF-like repeat 7"
misc_feature 4766. .4876
/note="Region: EGF-like repeat 8"
misc_feature 4976. .5119
/note="Region: CTC knot"
BASE COUNT 1393 a 1253 c 1264 g 1322 t 1 others
ORIGIN

Query Match 96.4%; Score 4588; DB 31; Length 5233;
Best Local Similarity 99.7%; Pred. No. 0.00e+00;
Matches 4674; Conservative 0; Mismatches 2; Indels 14; Gaps 3;

Db 557 ATGCGCGGCGTTGGCTGGCAGATGCTGTCCCTGTCGCTGGGGTTAGTGCTGGCGATCCTG 616
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1 atgcgcggcgttggctggcagatgctgtccctgtcgctggggttagtgctggcgatcctg 60

Db 617 AACAAGGTGGCACCGCAGGCGTGCCCGGCGCAGTGCTCTTGCTCGGGCAGCACAGTGGAC 676
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 61 aacaaggtggcaccgcaggcgtgcccggcgcagtgctcttgctcgggcagcacagtggac 120

Db 677 TGTCACGGGCTGGCGCTGCGCAGCGTGCCCAGGAATATCCCCCGCAACACCGAGAGACTG 736
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 tgtcacgggctggcgctgcgcagcgtgcccaggaatatcccccgcaacaccgagagactg 180

Db 737 GATTTAAATGGAAATAACATCACAAGAATTACGAAGACAGATTTTGCTGGTCTTAGACAT 796
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181 gatttaaatggaaataacatcacaagaattacgaagacagattttgctggtcttagacat 240

Db 797 CTAAGAGTTCTTCAGCTTATGGAGAATAAGATTAGCACCATTGAAAGAGGAGCATTCCAG 856
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 241 ctaagagttcttcagcttatggagaataagattagcaccattgaaagaggagcattccag 300

Db 857 GATCTTAAAGAACTAGAGAGACTGCGTTTAAACAGAAATCACCTTCAGCTGTTTCCTGAG 916
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 301 gatcttaaagaactagagagactgcgtttaaacagaaatcaccttcagctgtttcctgag 360

```
Db    917 TTGCTGTTTCTTGGGACTGCGAAGCTATACAGGCTTGATCTCAGTGAAAACCAAATTCAG 976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 ttgctgtttcttgggactgcgaagctatacaggcttgatctcagtgaaaaccaaattcag 420

Db    977 GCAATCCCAAGGAAAGCTTTCCGTGGGGCAGTTGACATAAAAAATTTGCAACTGGATTAC 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 gcaatcccaaggaaagctttccgtggggcagttgacataaaaaatttgcaactggattac 480

Db   1037 AACCAGATCAGCTGTATTGAAGATGGGGCATTCAGGGCTCTCCGGGACCTGGAAGTGCTC 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 aaccagatcagctgtattgaagatggggcattcagggctctccgggacctggaagtgctc 540

Db   1097 ACTCTCAACAATAACAACATTACTAGACTTTCTGTGGCAAGTTTCAACCATATGCCTAAA 1156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 actctcaacaataacaacattactagactttctgtggcaagtttcaaccatatgcctaaa 600

Db   1157 CTTAGGACTTTTCGACTGCATTCAAACAACCTGTATTGTGACTGCCACCTGGCCTGGCTC 1216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 cttaggacttttcgactgcattcaaacaacctgtattgtgactgccacctggcctggctc 660

Db   1217 TCCGACTGGCTTCGCCAAAGGCCTCGGGTTGGTCTGTACACTCAGTGTATGGGCCCCTCC 1276
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 tccgactggcttcgcaaaaggcctcgggttggtctgtacactcagtgtatgggcccctcc 720

Db   1277 CACCTGAGAGGCCATAATGTAGCCGAGGTTCAAAAACGAGAATTTGTCTGCAGTG----- 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 cacctgagaggccataatgtagccgaggttcaaaaacgagaatttgtctgcagtgatgag 780

Db   1332 -------GTCACCAGTCATTTATGGCTCCTTCTTGTAGTGTTTTGCACTGCCCTGCCGCC 1384
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 gaagaaggtcaccagtcatttatggctccttcttgtagtgttttgcactgccctgccgcc 840

Db   1385 TGTACCTGTAGCAACAATATCGTAGACTGTCGTGGGAAAGGTCTCACTGAGATCCCCACA 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 tgtacctgtagcaacaatatcgtagactgtcgtgggaaaggtctcactgagatccccaca 900

Db   1445 AATCTTCCAGAGACCATCACAGAAATACGTTTGGAACAGAACACAATCAAAGTCATCCCT 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 aatcttccagagaccatcacagaaatacgtttggaacagaacacaatcaaagtcatccct 960

Db   1505 CCTGGAGCTTTCTCACCATATAAAAAGCTTAGACGAATTGACCTGAGCAATAATCAGATC 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 cctggagctttctcaccatataaaaagcttagacgaattgacctgagcaataatcagatc 1020

Db   1565 TCTGAACTTGCACCAGATGCTTTCCAAGGACTACGCTCTCTGAATTCACTTGTCCTCTAT 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 tctgaacttgcaccagatgctttccaaggactacgctctctgaattcacttgtcctctat 1080

Db   1625 GGAAATAAAATCACAGAACTCCCCAAAAGTTTATTTGAAGGACTGTTTTCCTTACAGCTC 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 ggaaataaaatcacagaactccccaaaagtttatttgaaggactgttttccttacagctc 1140

Db   1685 CTATTATTGAATGCCAACAAGATAAACTGCCTTCGGGTAGATGCTTTTCAGGATCTCCAC 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 ctattattgaatgccaacaagataaactgccttcgggtagatgcttttcaggatctccac 1200

Db   1745 AACTTGAACCTTCTCTCCCTATATGACAACAAGCTTCAGACCATCGCCAAGGGGACCTTT 1804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 aacttgaaccttctctccctatatgacaacaagcttcagaccatcgccaaggggacctttt 1260

Db   1805 TCACCTCTTCGGGCCATTCAAACTATGCATTTGGCCCAGAACCCCTTTATTTGTGACTGC 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 tcacctcttcgggccattcaaactatgcatttggcccagaacccctttatttgtgactgc 1320

Db   1865 CATCTCAAGTGGCTAGCGGATTATCTCCATACCAACCCGATTGAGACCAGTGGTGCCCGT 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 catctcaagtggctagcggattatctccataccaacccgattgagaccagtggtgcccgt 1380

Db   1925 TGCACCAGCCCCCGCCGCCTGGCAAACAAAAGAATTGGACAGATCAAAAGCAAGAAATTC 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 tgcaccagccccgccgcctggcaaacaaaagaattggacagatcaaaagcaagaaattc 1440

Db   1985 CGTTGTTCAGGTACAGAAGATTATCGATCAAAATTAAGTGGAGACTGCTTTGCGGATCTG 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 cgttgttcaggtacagaagattatcgatcaaaattaagtggagactgctttgcggatctg 1500

Db   2045 GCTTGCCCTGAAAAGTGTCGCTGTGAAGGAACCACAGTAGATTGCTCTAATCAAAAGCTC 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 gcttgccctgaaaagtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagctc 1560

Db   2105 AACAAAATCCCGGAGCACATTCCCCAGTACACTGCAGAGTTGCGTCTCAATAATAATGAA 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 aacaaaatcccggagcacattccccagtacactgcagagttgcgtctcaataataatgaa 1620

Db   2165 TTTACCGTGTTGGAAGCCACAGGAATCTTTAAGAAACTTCCTCAATTACGTAAAATAAAC 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 tttaccgtgttggaagccacaggaatctttaagaaacttcctcaattacgtaaaataaac 1680

Db   2225 TTTAGCAACAATAAGATCACAGATATTGAGGAGGGAGCATTTGAAGGAGCATCTGGTGTA 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1681 tttagcaacaataagatcacagatattgaggagggagcatttgaaggagcatctggtgta 1740

Db   2285 AATGAAATACTTCTTACGAGTAATCGTTTGGAAAATGTGCAGCATAAGATGTTCAAGGGA 2344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 aatgaaatacttcttacgagtaatcgtttggaaaatgtgcagcataagatgttcaaggga 1800

Db   2345 TTGGAAAAGCCTCAAAACTT-GATGTTGAGAAGCAATCGAATAACCTGTGTGGGGAATGA 2403
          ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Qy   1801 ttggaaa-gcctcaaaactttgatgttgagaagcaatcgaataacctgtgtggggaatga 1859

Db   2404 CAGTTTCATAGGACTCAGTTCTGTGCGTATGCTTTCTTTGTATGATAATCAAATTACTAC 2463
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Qy   1860 cagtttcataggactcagttctgtgcgtttgctttctttgtatgataatcaaattactac 1919

Db   2464 AGTTGCACCAGGGGCATTTGATACTCTCCATTCTTTATCTACTCTAAACCTCTTGGCCAA 2523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1920 agttgcaccaggggcatttgatactctccattctttatctactctaaacctcttggccaa 1979

Db   2524 TCCTTTTAACTGTAACTGCTACCTGGCTTGGTTGGGAGAGTGGCTGAGAAAGAAGAGAAT 2583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1980 tcctttaactgtaactgctacctggcttggttgggagagtggctgagaaagaagagaat 2039

Db   2584 TGTCACGGGAAATCCTAGATGTCAAAAACCATACTTCCTGAAAGAAATACCCATCCAGGA 2643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2040 tgtcacgggaaatcctagatgtcaaaaaccatacttcctgaaagaaatacccatccagga 2099

Db   2644 TGTGGCCATTCAGGACTTCACTTGTGATGACGGAAATGATGACAATAGTTGCTCCCCACT 2703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2100 tgtggccattcaggacttcacttgtgatgacggaaatgatgacaatagttgctccccact 2159

Db   2704 TTCTCGCTGTCCTACTGAATGTACTTGCTTGGATACAGTCGTCCGATGTAGCAACAAGGG 2763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2160 ttctcgctgtcctactgaatgtacttgcttggatacagtcgtccgatgtagcaacaaggg 2219

Db   2764 TTTGAAGGTCTTGCCGAAAGGTATTCCAAGAGATGTCACAGAGTTGTATCTGGATGGAAA 2823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2220 tttgaaggtcttgccgaaaggtattccaagagatgtcacagagttgtatctggatggaaa 2279

Db   2824 CCAATTTACACTGGTTCCCAAGGAACTCTCCAACTACAAACATTTAACACTTATAGACTT 2883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2280 ccaatttacactggttcccaaggaactctccaactacaaacatttaacacttatagactt 2339

Db   2884 AAGTAACAACAGAATAAGCACGCTTTCTAATCAGAGCTTCAGCAACATGACCCAGCTCCT 2943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2340 aagtaacaacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcct 2399

Db   2944 CACCTTAATTCTTAGTTACAACCGTCTGAGATGTATTCCTCCTCGCACCTTTGATGGATT 3003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2400 caccttaattcttagttacaaccgtctgagatgtattcctcctcgcaccttgatggatt 2459

Db   3004 AAAGTCTCTTCGATTACTTTCTCTACATGGAAATGACATTTCTGTTGTGCCTGAAGGTGC 3063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2460 aaagtctcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgc 2519

Db   3064 TTTCAATGATCTTTCTGCATTATCACATCTAGCAATTGGAGCCAACCCTCTTTACTGTGA 3123
```

```
Qy  2520 tttcaatgatctttctgcattatcacatctagcaattggagccaaccctctttactgtga 2579

Db  3124 TTGTAACATGCAGTGGTTATCCGACTGGGTGAAGTCGGAATATAAGGAGCCTGGAATTGC 3183
Qy  2580 ttgtaacatgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgc 2639

Db  3184 TCGTTGTGCTGGTCCTGGAGAAATGGCAGATAAACTTTTACTCACAACTCCCTCCAAAAA 3243
Qy  2640 tcgttgtgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaa 2699

Db  3244 ATTTACCTGTCAAGGTCCTGTGGATGTCAATATTCTAGCTAAGTGTAACCCCTGCCTATC 3303
Qy  2700 atttacctgtcaaggtcctgtggatgtcaatattctagctaagtgtaaccccctgcctatc 2759

Db  3304 AAATCCGTGTAAAAATGATGGCACATGTAATAGTGATCCAGTTGACTTTTACCGATGCAC 3363
Qy  2760 aaatccgtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcac 2819

Db  3364 CTGTCCATATGGTTTCAAGGGGCAGGACTGTGATGTCCCAATTCATGCCTGCATCAGTAA 3423
Qy  2820 ctgtccatatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaa 2879

Db  3424 CCCATGTAAACATGGAGGAACTTGCCACTTAAAGGAAGGAGAAGAAGATGGATTCTGGTG 3483
Qy  2880 cccatgtaaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtg 2939

Db  3484 TATTTGTGCTGATGGATTTGAAGGAGAAAATTGTGAAGTCAACGTTGATGATTGTGAAGA 3543
Qy  2940 tatttgtgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaaga 2999

Db  3544 TAATGACTGTGAAAATAATTCTACATGTGTCGATGGCATTAATAACTACACATGCCTTTG 3603
Qy  3000 taatgactgtgaaaataattctacatgtgtcgatggcattaataactacacatgcctttg 3059

Db  3604 CCCACCTGAGTATACAGGTGAGTTGTGTGAGGAGAAGCTGGACTTCTGTGCCCAGGACCT 3663
Qy  3060 cccacctgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacct 3119

Db  3664 GAACCCCTGCCAGCACGATTCAAAGTGCATCCTAACTCCAAAGGGATTCAAATGTGACTG 3723
Qy  3120 gaacccctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactg 3179

Db  3724 CACACCAGGGTACGTAGGTGAACACTGCGACATCGATTTTGACGACTGCCAAGACAACAA 3783
Qy  3180 cacaccagggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaa 3239

Db  3784 GTGTAAAAACGGAGCCCACTGCACAGATGCAGTGAACGGCTATACGTGCATATGCCCCGA 3843
Qy  3240 gtgtaaaaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccga 3299

Db  3844 AGGTTACAGTGGCTTGTTCTGTGAGTTTTCTCCACCCATGGTCCTCCCTCGTACCAGCCC 3903
Qy  3300 aggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagccc 3359

Db  3904 CTGTGATAATTTTGATTGTCAGAATGGAGCTCAGTGTATCGTCAGAATAAATGAGCCAAT 3963
Qy  3360 ctgtgataattttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaat 3419

Db  3964 ATGTCAGTGTTTGCCTGGCTATCAGGGAGAAAAGTGTGAAAAATTGGTTAGTGTGAATTT 4023
Qy  3420 atgtcagtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaattt 3479

Db  4024 TATAAACAAAGAGTCTTATCTTCAGATTCCTTCAGCCAAGGTTCGGCCTCAGACGAACAT 4083
Qy  3480 tataaacaaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacat 3539

Db  4084 AACACTTCAGATTGCCACAGATGAAGACAGCGGAATCCTCCTGTATAAGGGTGACAAAGA 4143
Qy  3540 aacacttcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaaga 3599

Db  4144 CCATATCGCGGTAGAACTCTATCGGGGGCGTGTTCGTGCCAGCTATGACACCGGCTCTCA 4203

Qy  3600 ccatatcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctca 3659

Db  4204 TCCAGCTTCTGCCATTTACAGTGTGGAGACAATCAATGATGGAAACTTCCACATTGTGGA 4263
Qy  3660 tccagcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtgga 3719

Db  4264 ACTACTTGCCTTGGATCAGAGTCTCTCTTTGTCCGTGGATGGTGGGAACCCCAAAATCAT 4323
Qy  3720 actacttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcat 3779

Db  4324 CACTAACTTGTCAAAGCAGTCCACTCTGAATTTTGACTCTCCACTCTATGTAGGAGGCAT 4383
Qy  3780 cactaacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcat 3839

Db  4384 GCCAGGGAAGAGTAACGTGGCATCTCTGCGCCAGGCCCCTGGGCAGAACGGAACCAGCTT 4443
Qy  3840 gccagggaagagtaacgtggcatctctgcgccaggccctgggcagaacggaaccagctt 3899

Db  4444 CCACGGCTGCATCCGGAACCTTTACATCAACAGTGAGCTGCAGGACTTCCAGAAGGTGCC 4503
Qy  3900 ccacggctgcatccggaacctttacatcaacagtgagctgcaggacttccagaaggtgcc 3959

Db  4504 GATGCAAACAGGCATTTTGCCTGGCTGTGAGCCATGCCACAAGAAGGTGTGTGCCCATGG 4563
Qy  3960 gatgcaaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatgg 4019

Db  4564 CACATGCCAGCCCAGCAGCCAGGCAGGCTTCACCTGCGAGTGCCAGGAAGGATGGATGGG 4623
Qy  4020 cacatgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggg 4079

Db  4624 GCCCCTCTGTGACCAACGGACCAATGACCCTTGCCTTGGAAATAAATGCGTACATGGCAC 4683
Qy  4080 gcccctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcac 4139

Db  4684 CTGCTTGCCCATCAATGCGTTCTCCTACAGCTGTAAGTGCTTGGAGGGCCATGGAGGTGT 4743
Qy  4140 ctgcttgcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgt 4199

Db  4744 CCTCTGTGATGAAGAGGAGGATCTGTTTAACCCATGCCAGGCGATCAAGTGCAAGCATGG 4803
Qy  4200 cctctgtgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatgg 4259

Db  4804 GAAGTGCAGGCTTTCAGGTCTGGGGCAGCCCTACTGTGAATGCAGCAGTGGATACACGGG 4863
Qy  4260 gaagtgcaggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacggg 4319

Db  4864 GGACAGCTGTGATCGAGAAATCTCTTGTCGAGGGGAAAGGATAAGAGATTATTACCAAAA 4923
Qy  4320 ggacagctgtgatcgagaaatctcttgtcgaggggaaaggataagagattattaccaaaa 4379

Db  4924 GCAGCAGGGCTATGCTGCTTGCCAAACAACCAAGAAGGTGTCCCGATTAGAGTGCAGAGG 4983
Qy  4380 gcagcagggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagagg 4439

Db  4984 TGGGTGTGCAGGAGGGCAGTGCTGTGGACCGCTGAGGAGCAAGCGGCGGAAATACTCTTT 5043
Qy  4440 tgggtgtgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactcttt 4499

Db  5044 CGAATGCACTGACGGCTCCTCCTTTGTGGACGAGGTTGAGAAAGTGGTGAAGTGCGGCTG 5103
Qy  4500 cgaatgcactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559

Db  5104 TACGAGGTGTGTGTCCTAAACACACTCCCGGCAGCTCTGTCTTTGGAAAAGGTTGTATAC 5163
Qy  4560 tacgaggtgtgtgtcctaaacacactcccggcagctctgtctttggaaaaggttgtatac 4619

Db  5164 TTCTTGACCATGTGGGACTAATGAATGCTTCATAGTGGAAATATTTGAAATATATTGTAA 5223
Qy  4620 ttcttgaccatgtgggactaatgaatgcttcatagtggaaatatttgaaatatattgtaa 4679

Db  5224 AATACAGAAC 5233
Qy  4680 aatacagaac 4689
```

```
RESULT    2
LOCUS       AB017168     4950 bp    mRNA            PRI       06-FEB-1999
DEFINITION  Homo sapiens mRNA for Slit-2 protein, complete cds.
ACCESSION   AB017168
NID         g4049586
VERSION     AB017168.1  GI:4049586
KEYWORDS    slit-2; Slit-2 protein.
SOURCE      Homo sapiens fetal tissue_lib:lung cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4950)
  AUTHORS   Itoh,A. and Sakano,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1998) to the DDBJ/EMBL/GenBank databases. Akira
            Itoh, Asahi Chemical Industry co.,ltd., Life Science Fundamental
            Research Laboratory; 2-1, Samejima, Fuji, Shizuoka 416-8501, Japan
            (E-mail:a8611483@ut.asahi-kasei.co.jp, Tel:+81-545-62-3231,
            Fax:+81-545-62-3249)
REFERENCE   2  (sites)
  AUTHORS   Itoh,A., Miyabayashi,T., Ohno,M. and Sakano,S.
  TITLE     Cloning and expressions of three mammalian homologues of Drosophila
            slit suggest possible roles for Slit in the formation and
            maintenance of the nervous system
  JOURNAL   Brain Res. Mol. Brain Res. 62 (2), 175-186 (1998)
  MEDLINE   99033071
FEATURES             Location/Qualifiers
     source          1. .4950
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="fetal"
                     /tissue_lib="lung"
     gene            205. .4794
                     /gene="slit-2"
     CDS             205. .4794
                     /gene="slit-2"
                     /codon_start=1
                     /product="Slit-2 protein"
                     /protein_id="BAA35185.1"
                     /db_xref="PID:d1036171"
                     /db_xref="PID:g4049587"
                     /db_xref="GI:4049587"
                     /translation="MRGVGWQMLSLSLGLVLAILNKVAPQACPAQCSCSGSTVDCHGL
                     ALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDL
                     KELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIPRKAFRGAVDIKNLQLDY
                     NQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRTFRLHSNNLYCDCHLA
                     WLSDWLRQRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSGHQSFMAPSCSVLHCPAA
                     CTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNN
                     QISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLLLNANKINCLRVDAF
                     QDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHLKWLADYLHTNPI
                     ETSGARCTSPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYRSKLSGDCFADLACPE
                     KCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRKINFS
                     NNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRITCVGND
                     SFIGLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCYLAWLGEWLRKK
                     RIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPTECTCLDTVVRC
                     SNKGLKVLPKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQSFS
                     NMTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAI
                     GANPLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVN
                     ILAKCNPCLSNPCKNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACISNPCKHGGTC
                     HLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTG
                     ELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDCQDNKCKNG
                     AHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRINEPICQ
                     CLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKD
                     HIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSVDGGNPK
                     IITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDF
                     QKVPMQTGILPGCEPCHKKVCAHGTCQPSSQAGFTCECQEGWMGPLCDQRTNDPCLGN
                     KCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQAIKCKHGKCRLSGLGQPYC
                     ECSSGYTGDSCDREISCRGERIRDYYQKQQGYAACQTTKKVSRLECRGGCAGGQCCGP
                     LRSKRRKYSFECTDGSSFVDEVEKVVKCGCTRCVS"
BASE COUNT     1355 a   1138 c   1186 g   1271 t
ORIGIN
```

```
Query Match          94.6%;  Score 4502;  DB 29;  Length 4950;
Best Local Similarity  99.2%;  Pred. No. 0.00e+00;
Matches  4720;  Conservative    0;  Mismatches   2;  Indels  36;  Gaps   2;

Db    205 ATGCGCGGCGTTGGCTGGCAGATGCTGTCCCTGTCGCTGGGGTTAGTGCTGGCGATCCTG 264
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 atgcgcggcgttggctggcagatgctgtccctgtcgctggggttagtgctggcgatcctg 60

Db    265 AACAAGGTGGCACCGCAGGCGTGCCCGGCGCAGTGCTCTTGCTCGGGCAGCACAGTGGAC 324
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 aacaaggtggcaccgcaggcgtgcccggcgcagtgctcttgctcgggcagcacagtggac 120

Db    325 TGTCACGGGCTGGCGCTGCGCAGCGTGCCCAGGAATATCCCCCGCAACACCGAGAGACTG 384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 tgtcacgggctggcgctgcgcagcgtgcccaggaatatccccgcaacaccgagagactg 180

Db    385 GATTTAAATGGAAATAACATCACAAGAATTACGAAGACAGATTTTGCTGGTCTTAGACAT 444
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 gatttaaatggaaataacatcacaagaattacgaagacagattttgctggtcttagacat 240

Db    445 CTAAGAGTTCTTCAGCTTATGGAGAATAAGATTAGCACCATTGAAAGAGGAGCATTCCAG 504
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 ctaagagttcttcagcttatggagaataagattagcaccattgaaagaggagcattccag 300

Db    505 GATCTTAAAGAACTAGAGAGACTGCGTTTAAACAGAAATCACCTTCAGCTGTTTCCTGAG 564
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 gatcttaaagaactagagagactgcgtttaaacagaaatcaccttcagctgtttcctgag 360

Db    565 TTGCTGTTTCTTGGGACTGCGAAGCTATACAGGCTTGATCTCAGTGAAAACCAAATTCAG 624
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 ttgctgtttcttgggactgcgaagctatacaggcttgatctcagtgaaaaccaaattcag 420

Db    625 GCAATCCCAAGGAAAGCTTTCCGTGGGGCAGTTGACATAAAAAATTTGCAACTGGATTAC 684
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 gcaatcccaaggaaagctttccgtggggcagttgacataaaaaatttgcaactggattac 480

Db    685 AACCAGATCAGCTGTATTGAAGATGGGGCATTCAGGGCTCTCCGGGACCTGGAAGTGCTC 744
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 aaccagatcagctgtattgaagatggggcattcagggctctccgggacctggaagtgctc 540

Db    745 ACTCTCAACAATAACAACATTACTAGACTTTCTGTGGCAAGTTTCAACCATATGCCTAAA 804
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 actctcaacaataacaacattactagactttctgtggcaagtttcaaccatatgcctaaa 600

Db    805 CTTAGGACTTTTCGACTGCATTCAAACAACCTGTATTGTGACTGCCACCTGGCCTGGCTC 864
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 cttaggacttttcgactgcattcaaacaacctgtattgtgactgccacctggcctggctc 660

Db    865 TCCGACTGGCTTCGCCAAAGGCCTCGGGTTGGTCTGTACACTCAGTGTATGGGCCCCTCC 924
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 tccgactggcttcgcaaaaggcctcgggttggtctgtacactcagtgtatgggcccctcc 720

Db    925 CACCTGAGAGGCCATAATGTAGCCGAGGTTCAAAAACGAGAATTTGTCTGCAGTG----- 979
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 cacctgagaggccataatgtagccgaggttcaaaaacgagaatttgtctgcagtgatgag 780

Db    980 -------GTCACCAGTCATTTATGGCTCCTTCTTGTAGTGTTTTGCACTGCCCTGCCGCC 1032
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 gaagaaggtcaccagtcatttatggctccttcttgtagtgttttgcactgccctgccgcc 840

Db   1033 TGTACCTGTAGCAACAATATCGTAGACTGTCGTGGGAAAGGTCTCACTGAGATCCCCACA 1092
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 tgtacctgtagcaacaatatcgtagactgtcgtgggaaaggtctcactgagatccccaca 900

Db   1093 AATCTTCCAGAGACCATCACAGAAATACGTTTGGAACAGAACACAATCAAAGTCATCCCT 1152
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 aatcttccagagaccatcacagaaatacgtttggaacagaacacaatcaaagtcatccct 960

Db   1153 CCTGGAGCTTTCTCACCATATAAAAAGCTTAGACGAATTGACCTGAGCAATAATCAGATC 1212
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 cctggagctttctcaccatataaaaagcttagacgaattgacctgagcaataatcagatc 1020
```

```
Db  1213 TCTGAACTTGCACCAGATGCTTTCCAAGGACTACGCTCTCTGAATTCACTTGTCCTCTAT 1272
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1021 tctgaacttgcaccagatgctttccaaggactacgctctctgaattcacttgtcctctat 1080

Db  1273 GGAAATAAAATCACAGAACTCCCCAAAAGTTTATTTGAAGGACTGTTTTCCTTACAGCTC 1332
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1081 ggaaataaaatcacagaactccccaaaagtttatttgaaggactgttttccttacagctc 1140

Db  1333 CTATTATTGAATGCCAACAAGATAAACTGCCTTCGGGTAGATGCTTTTCAGGATCTCCAC 1392
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1141 ctattattgaatgccaacaagataaactgccttcgggtagatgcttttcaggatctccac 1200

Db  1393 AACTTGAACCTTCTCTCCCTATATGACAACAAGCTTCAGACCATCGCCAAGGGGACCTTT 1452
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1201 aacttgaaccttctctccctatatgacaacaagcttcagaccatcgccaaggggaccttt 1260

Db  1453 TCACCTCTTCGGGCCATTCAAACTATGCATTTGGCCCAGAACCCCTTTATTTGTGACTGC 1512
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1261 tcacctcttcgggccattcaaactatgcatttggcccagaaccccttttatttgtgactgc 1320

Db  1513 CATCTCAAGTGGCTAGCGGATTATCTCCATACCAACCCGATTGAGACCAGTGGTGCCCGT 1572
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1321 catctcaagtggctagcggattatctccataccaacccgattgagaccagtggtgcccgt 1380

Db  1573 TGCACCAGCCCCCGCCGCCTGGCAAACAAAAGAATTGGACAGATCAAAAGCAAGAAATTC 1632
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1381 tgcaccagcccccgccgcctggcaaacaaaagaattggacagatcaaaagcaagaaattc 1440

Db  1633 CGTTGTTCAGCTAAAGAACAGTATTTCATTCCAGGTACAGAAGATTATCGATCAAAATTA 1692
         ||||||||||                    ||||||||||||||||||||||||||||
Qy  1441 cgttgttcag----------------------gtacagaagattatcgatcaaaatta 1476

Db  1693 AGTGGAGACTGCTTTGCGGATCTGGCTTGCCCTGAAAAGTGTCGCTGTGAAGGAACCACA 1752
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1477 agtggagactgctttgcggatctggcttgccctgaaaagtgtcgctgtgaaggaaccaca 1536

Db  1753 GTAGATTGCTCTAATCAAAAGCTCAACAAAATCCCGGAGCACATTCCCCAGTACACTGCA 1812
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1537 gtagattgctctaatcaaaagctcaacaaaatcccggagcacattccccagtacactgca 1596

Db  1813 GAGTTGCGTCTCAATAATAATGAATTTACCGTGTTGGAAGCCACAGGAATCTTTAAGAAA 1872
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1597 gagttgcgtctcaataataatgaatttaccgtgttggaagccacaggaatctttaagaaa 1656

Db  1873 CTTCCTCAATTACGTAAAATAAACTTTAGCAACAATAAGATCACAGATATTGAGGAGGGA 1932
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1657 cttcctcaattacgtaaaataaactttagcaacaataagatcacagatattgaggaggga 1716

Db  1933 GCATTTGAAGGAGCATCTGGTGTAAATGAAATACTTCTTACGAGTAATCGTTTGGAAAAT 1992
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1717 gcatttgaaggagcatctggtgtaaatgaaatacttcttacgagtaatcgtttggaaaat 1776

Db  1993 GTGCAGCATAAGATGTTCAAGGGATTGGAAAGCCTCAAAACTTTGATGTTGAGAAGCAAT 2052
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1777 gtgcagcataagatgttcaagggattggaaagcctcaaaactttgatgttgagaagcaat 1836

Db  2053 CGAATAACCTGTGTGGGGAATGACAGTTTCATAGGACTCAGTTCTGTGCGTTTGCTTTCT 2112
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1837 cgaataacctgtgtggggaatgacagtttcataggactcagttctgtgcgtttgctttct 1896

Db  2113 TTGTATGATAATCAAATTACTACAGTTGCACCAGGGGCATTTGATACTCTCCATTCTTTA 2172
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1897 ttgtatgataatcaaattactacagttgcaccaggggcatttgatactctccattcttta 1956

Db  2173 TCTACTCTAAACCTCTTGGCCAATCCTTTTAACTGTAACTGCTACCTGGCTTGGTTGGGA 2232
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1957 tctactctaaacctcttggccaatccttttaactgtaactgctacctggcttggttggga 2016

Db  2233 GAGTGGCTGAGAAAGAAGAGAATTGTCACGGGAAATCCTAGATGTCAAAAACCATACTTC 2292
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2017 gagtggctgagaaagaagagaattgtcacgggaaatcctagatgtcaaaaaccatacttc 2076

Db  2293 CTGAAAGAAATACCCATCCAGGATGTGGCCATTCAGGACTTCACTTGTGATGACGGAAAT 2352
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2077 ctgaaagaaatacccatccaggatgtggccattcaggacttcacttgtgatgacggaaat 2136

Db  2353 GATGACAATAGTTGCTCCCCACTTTCTCGCTGTCCTACTGAATGTACTTGCTTGGATACA 2412
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2137 gatgacaatagttgctccccactttctcgctgtcctactgaatgtacttgcttggataca 2196

Db  2413 GTCGTCCGATGTAGCAACAAGGGTTTGAAGGTCTTGCCGAAAGGTATTCCAAGAGATGTC 2472
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2197 gtcgtccgatgtagcaacaagggtttgaaggtcttgccgaaaggtattccaagagatgtc 2256

Db  2473 ACAGAGTTGTATCTGGATGGAAACCAATTTACACTGGTTCCCAAGGAACTCTCCAACTAC 2532
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2257 acagagttgtatctggatggaaaccaatttacactggttcccaaggaactctccaactac 2316

Db  2533 AAACATTTAACACTTATAGACTTAAGTAACAACAGAATAAGCACGCTTTCTAATCAGAGC 2592
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2317 aaacatttaacacttatagacttaagtaacaacagaataagcacgctttctaatcagagc 2376

Db  2593 TTCAGCAACATGACCCAGCTCCTCACCTTAATTCTTAGTTACAACCGTCTGAGATGTATT 2652
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2377 ttcagcaacatgacccagctcctcaccttaattcttagttacaaccgtctgagatgtatt 2436

Db  2653 CCTCCTCGCACCTTTGATGGATTAAAGTCTCTTCGATTACTTTCTCTACATGGAAATGAC 2712
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2437 cctcctcgcaccttt gatggattaaagtctcttcgattactttctctacatggaaatgac 2496

Db  2713 ATTTCTGTTGTGCCTGAAGGTGCTTTCAATGATCTTTCTGCATTATCACATCTAGCAATT 2772
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2497 atttctgttgtgcctgaaggtgctttcaatgatctttctgcattatcacatctagcaatt 2556

Db  2773 GGAGCCAACCCTCTTTACTGTGATTGTAACATGCAGTGGTTATCCGACTGGGTGAAGTCG 2832
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2557 ggagccaaccctctttactgtgattgtaacatgcagtggttatccgactgggtgaagtcg 2616

Db  2833 GAATATAAGGAGCCTGGAATTGCTCGTTGTGCTGGTCCTGGAGAAATGGCAGATAAACTT 2892
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2617 gaatataaggagcctggaattgctcgttgtgctggtcctggagaaatggcagataaactt 2676

Db  2893 TTACTCACAACTCCCTCCAAAAAATTTACCTGTCAAGGTCCTGTGGATGTCAATATTCTA 2952
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2677 ttactcacaactccctccaaaaaatttacctgtcaaggtcctgtggatgtcaatattcta 2736

Db  2953 GCTAAGTGTAACCCCTGCCTATCAAATCCGTGTAAAAATGATGGCACATGTAATAGTGAT 3012
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2737 gctaagtgtaaccccctgcctatcaaatccgtgtaaaaatgatggcacatgtaatagtgat 2796

Db  3013 CCAGTTGACTTTTACCGATGCACCTGTCCATATGGTTTCAAGGGGCAGGACTGTGATGTC 3072
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2797 ccagttgacttttaccgatgcacctgtccatatggtttcaaggggcaggactgtgatgtc 2856

Db  3073 CCAATTCATGCCTGCATCAGTAACCCATGTAAACATGGAGGAACTTGCCACTTAAAGGAA 3132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2857 ccaattcatgcctgcatcagtaacccatgtaaacatggaggaacttgccacttaaaggaa 2916

Db  3133 GGAGAAGAAGATGGATTCTGGTGTATTTGTGCTGATGGATTTGAAGGAGAAAATTGTGAA 3192
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2917 ggagaagaagatggattctggtgtatttgtgctgatggatttgaaggagaaaattgtgaa 2976

Db  3193 GTCAACGTTGATGATTGTGAAGATAATGACTGTGAAAATAATTCTACATGTGTCGATGGC 3252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2977 gtcaacgttgatgattgtgaagataatgactgtgaaaataattctacatgtgtcgatggc 3036

Db  3253 ATTAATAACTACACATGCCTTTGCCCACCTGAGTATACAGGTGAGTTGTGTGAGGAGAAG 3312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3037 attaataactacacatgcctttgcccacctgagtatacaggtgagttgtgtgaggagaag 3096

Db  3313 CTGGACTTCTGTGCCCAGGACCTGAACCCCTGCCAGCACGATTCAAAGTGCATCCTAACT 3372
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3097 ctggacttctgtgcccaggacctgaaccctgccagcacgattcaaagtgcatcctaact 3156

Db  3373 CCAAAGGGATTCAAATGTGACTGCACACCAGGGTACGTAGGTGAACACTGCGACATCGAT 3432
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   3157 ccaaagggattcaaatgtgactgcacaccagggtacgtaggtgaacactgcgacatcgat 3216

Db   3433 TTTGACGACTGCCAAGACAACAAGTGTAAAAACGGAGCCCACTGCACAGATGCAGTGAAC 3492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3217 tttgacgactgccaagacaacaagtgtaaaaacggagcccactgcacagatgcagtgaac 3276

Db   3493 GGCTATACGTGCATATGCCCCGAAGGTTACAGTGGCTTGTTCTGTGAGTTTTCTCCACCC 3552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3277 ggctatacgtgcatatgccccgaaggttacagtggcttgttctgtgagttttctccaccc 3336

Db   3553 ATGGTCCTCCCTCGTACCAGCCCCTGTGATAATTTTGATTGTCAGAATGGAGCTCAGTGT 3612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3337 atggtcctccctcgtaccagcccctgtgataattttgattgtcagaatggagctcagtgt 3396

Db   3613 ATCGTCAGAATAAATGAGCCAATATGTCAGTGTTTGCCTGGCTATCAGGGAGAAAAGTGT 3672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3397 atcgtcagaataaatgagccaatatgtcagtgtttgcctggctatcagggagaaaagtgt 3456

Db   3673 GAAAAATTGGTTAGTGTGAATTTTATAAACAAAGAGTCTTATCTTCAGATTCCTTCAGCC 3732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3457 gaaaaattggttagtgtgaattttataaacaaagagtcttatcttcagattccttcagcc 3516

Db   3733 AAGGTTCGGCCTCAGACGAACATAACACTTCAGATTGCCACAGATGAAGACAGCGGAATC 3792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3517 aaggttcggcctcagacgaacataacacttcagattgccacagatgaagacagcggaatc 3576

Db   3793 CTCCTGTATAAGGGTGACAAAGACCATATCGCGGTAGAACTCTATCGGGGGCGTGTTCGT 3852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3577 ctcctgtataagggtgacaaagaccatatcgcggtagaactctatcgggggcgtgttcgt 3636

Db   3853 GCCAGCTATGACACCGGCTCTCATCCAGCTTCTGCCATTTACAGTGTGGAGACAATCAAT 3912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3637 gccagctatgacaccggctctcatccagcttctgccatttacagtgtggagacaatcaat 3696

Db   3913 GATGGAAACTTCCACATTGTGGAACTACTTGCCTTGGATCAGAGTCTCTCTTTGTCCGTG 3972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3697 gatggaaacttccacattgtggaactacttgccttggatcagagtctctctttgtccgtg 3756

Db   3973 GATGGTGGGAACCCCAAAATCATCACTAACTTGTCAAAGCAGTCCACTCTGAATTTTGAC 4032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3757 gatggtgggaaccccaaaatcatcactaacttgtcaaagcagtccactctgaattttgac 3816

Db   4033 TCTCCACTCTATGTAGGAGGCATGCCAGGGAAGAGTAACGTGGCATCTCTGCGCCAGGCC 4092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3817 tctccactctatgtaggaggcatgccagggaagagtaacgtggcatctctgcgccaggcc 3876

Db   4093 CCTGGGCAGAACGGAACCAGCTTCCACGGCTGCATCCGGAACCTTTACATCAACAGTGAG 4152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3877 cctgggcagaacggaaccagcttccacggctgcatccggaacctttacatcaacagtgag 3936

Db   4153 CTGCAGGACTTCCAGAAGGTGCCGATGCAAACAGGCATTTTGCCTGGCTGTGAGCCATGC 4212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3937 ctgcaggacttccagaaggtgccgatgcaaacaggcattttgcctggctgtgagccatgc 3996

Db   4213 CACAAGAAGGTGTGTGCCCATGGCACATGCCAGCCCAGCAGCCAGGCAGGCTTCACCTGC 4272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3997 cacaagaaggtgtgtgcccatggcacatgccagcccagcagccaggcaggcttcacctgc 4056

Db   4273 GAGTGCCAGGAAGGATGGATGGGGCCCCTCTGTGACCAACGGACCAATGACCCTTGCCTT 4332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4057 gagtgccaggaaggatggatggggcccctctgtgaccaacggaccaatgacccttgcctt 4116

Db   4333 GGAAATAAATGCGTACATGGCACCTGCTTGCCCATCAATGCGTTCTCCTACAGCTGTAAG 4392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4117 ggaaataaatgcgtacatggcacctgcttgcccatcaatgcgttctcctacagctgtaag 4176

Db   4393 TGCTTGGAGGGCCATGGAGGTGTCCTCTGTGATGAAGAGGAGGATCTGTTTAACCCATGC 4452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4177 tgcttggagggccatggaggtgtcctctgtgatgaagaggaggatctgtttaacccatgc 4236

Db   4453 CAGGCGATCAAGTGCAAGCACGGGAAGTGCAGGCTTTCAGGTCTGGGGCAGCCCTACTGT 4512
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Qy   4237 caggcgatcaagtgcaagcatgggaagtgcaggctttcaggtctggggcagccctactgt 4296

Db   4513 GAATGCAGCAGTGGATACACGGGGGACAGCTGTGATCGAGAAATCTCTTGTCGAGGGGAA 4572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4297 gaatgcagcagtggatacacgggggacagctgtgatcgagaaatctcttgtcgaggggaa 4356

Db   4573 AGGATAAGAGATTATTACCAAAAGCAGCAGGGCTATGCTGCTTGCCAAACAACCAAGAAG 4632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4357 aggataagagattattaccaaaagcagcagggctatgctgcttgccaaacaaccaagaag 4416

Db   4633 GTGTCCCGATTAGAGTGCAGAGGTGGGTGTGCAGGAGGGCAGTGCTGTGGACCGCTGAGG 4692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4417 gtgtcccgattagagtgcagaggtgggtgtgcaggagggcagtgctgtggaccgctgagg 4476

Db   4693 AGCAAGCGGCGGAAATACTCTTTCGAATGCACTGACGGCTCCTCCTTTGTGGACGAGGTT 4752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4477 agcaagcggcggaaatactctttcgaatgcactgacggctcctcctttgtggacgaggtt 4536

Db   4753 GAGAAAGTGGTGAAGTGCGGCTGTACGAGGTGTGTGTCCTAAACACACTCCCGGCAGCTC 4812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4537 gagaaagtggtgaagtgcggctgtacgaggtgtgtgtcctaaacacactcccggcagctc 4596

Db   4813 TGTCTTTGGAAAAGGTTGTATACTTCTTGACCATGTGGGACTAATGAATGCTTCATAGTG 4872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4597 tgtctttggaaaaggttgtatacttcttgaccatgtgggactaatgaatgcttcatagtg 4656

Db   4873 GAAATATTTGAAATATATTGTAAAATACAGAACAGACTTATTTTTATTATGAGAATAAAG 4932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4657 gaaatatttgaaatatattgtaaaatacagaacagacttatttttattatgagaataaag 4716

Db   4933 ACTTTTTTTCTGCATTTG 4950
          |||||||||||||||||
Qy   4717 actttttttctgcatttg 4734
```

```
RESULT    3
LOCUS       AF074960     5121 bp    mRNA            ROD       04-MAR-1999
DEFINITION  Mus musculus neurogenic extracellular slit protein (Slit2) mRNA,
            partial cds.
ACCESSION   AF074960
NID         g4151258
VERSION     AF074960.1  GI:4151258
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5121)
  AUTHORS   Holmes,G.P., Negus,K., Burridge,L., Raman,S., Algar,E., Yamada,T.
            and Little,M.H.
  TITLE     Distinct but overlapping expression patterns of two vertebrate slit
            homologs implies functional roles in CNS development and
            organogenesis
  JOURNAL   Mech. Dev. 79 (1-2), 57-72 (1998)
REFERENCE   2  (bases 1 to 5121)
  AUTHORS   Holmes,G.P., Negus,K., Smith,K., Yamada,T. and Little,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1998) Center for Molecular and Cellular Biology,
            University of Queensland, St. Lucia, Brisbane, Queensland 4072,
            Australia
FEATURES             Location/Qualifiers
     source          1. .5121
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1. .5121
                     /gene="Slit2"
     CDS             <1. .3078
                     /gene="Slit2"
                     /note="similar to Homo sapiens SLIT2; similar to
                     Drosophila slit family"
                     /codon_start=1
                     /product="neurogenic extracellular slit protein"
                     /protein_id="AAD04345.1"
                     /db_xref="PID:g4151259"
```

```
                /db_xref="GI:4151259"
                /translation="ACPEKCRCEGTTVDCSNQRLNKIPDHIPQYTAELRLNNNEFTVL
                EATGIFKKLPQLRXINFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLE
                SLKTLMLRSNRISCVGNDSFIGLGSVRLLSLYDNQITTVAPGAFDXLHSLSTLNLLAN
                PFNCNCHLAWLGEWLRRKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCS
                PLSRCPSECTCLDTXVRCSNKGLKVLPKGIPKDVTELYLDGNQFTLVPKELSNYKHLT
                LIDLSNNRISTLSNQXFSNMTQLLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDIS
                VVPEGAFNDLSALSHLAIGANPLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKL
                LLTTPSKKFTCQGPMDITIQAKCNPCLSNPCKNDGTCNNDPVDFYRCTCPYGFKGQDC
                DVPIHACISNPCKHGGTCHLKEGENAGFWCTCADGFEGENCEVNIDDCEDNDCENNST
                CVDGINNYTCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYIG
                EHCDIDFDDCQDNKCKNGAHCTDAVNGYTCVCPEGYSGLFCEFSPPMVLPRTSPCDNF
                DCQNGAQCIIRINEPICQCLPGYLGEKCEKLVSVNFVNKESYLQIPSAKVRPQTNITL
                QIATDEDSGILLYKGDKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVE
                LLTLDSSLSLSVDGGSPKVITNLSKQSTLNFDSPLYVGGMPGKNNVASLRQAPGQNGT
                SFHGCIRNLYINSELQDFRKMPMQTGILPGCEPCHKKVCAHGMCQPSSQSGFTCECEE
                GWMGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCKCLEGHGGVLCDEEEDLFNPCQM
                IKCKHGKCRLSGVGQPYCECNSGFTGDSCDREISCRGERIRDYYQKQQGYAACQTTKK
                VSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCARCAS"
misc_feature    4. .96
                /gene="Slit2"
                /note="Region: amino-flanking"
misc_feature    100. .132
                /gene="Slit2"
                /note="Region: leucine rich-region 3-1"
misc_feature    148. .207
                /gene="Slit2"
                /note="Region: leucine rich-region 3-1A"
misc_feature    220. .279
                /gene="Slit2"
                /note="Region: leucine rich-region 3-2"
misc_feature    292. .351
                /gene="Slit2"
                /note="Region: leucine rich-region 3-3"
misc_feature    364. .453
                /gene="Slit2"
                /note="Region: leucine rich-region 3-4"
misc_feature    436. .471
                /gene="Slit2"
                /note="Region: leucine rich-region 3-5"
misc_feature    478. .666
                /gene="Slit2"
                /note="Region: carboxy-flanking"
misc_feature    667. .759
                /gene="Slit2"
                /note="Region: amino-flanking"
misc_feature    763. .795
                /gene="Slit2"
                /note="Region: leucine rich region 4-1"
misc_feature    805. .864
                /gene="Slit2"
                /note="Region: leucine rich region 4-2"
misc_feature    877. .936
                /gene="Slit2"
                /note="Region: leucine rich region 4-3"
misc_feature    949. .1008
                /gene="Slit2"
                /note="Region: leucine rich region 4-4"
misc_feature    1021. .1056
                /gene="Slit2"
                /note="Region: leucine rich region 4-5"
misc_feature    1063. .1236
                /gene="Slit2"
                /note="Region: carboxy-flanking"
misc_feature    1240. .1353
                /gene="Slit2"
                /note="Region: EGF-like repeat 1"
misc_feature    1357. .1476
                /gene="Slit2"
                /note="Region: EGF-like repeat 2"
misc_feature    1480. .1590
                /gene="Slit2"
                /note="Region: EGF-like repeat 3"
misc_feature    1594. .1710
                /gene="Slit2"
                /note="Region: EGF-like repeat 4"
misc_feature    1714. .1824
                /gene="Slit2"
                /note="Region: EGF-like repeat 5"
misc_feature    1849. .1959
                /gene="Slit2"
                /note="Region: EGF-like repeat 6"
misc_feature    1960. .2475
                /gene="Slit2"
                /note="Region: agrin perlecan laminin slit motif"
misc_feature    2476. .2592
                /gene="Slit2"
                /note="Region: EGF-like repeat 6A"
misc_feature    2599. .2709
                /gene="Slit2"
                /note="Region: EGF-like repeat 7"
misc_feature    2722. .2832
                /gene="Slit2"
                /note="Region: EGF-like repeat 8"
misc_feature    2932. .3075
                /gene="Slit2"
                /note="Region: CTC knot"
BASE COUNT     1533 a   1060 c   1106 g   1418 t      4 others
ORIGIN

  Query Match          52.5%;  Score 2498;  DB 32;  Length 5121;
  Best Local Similarity  88.6%;  Pred. No. 0.00e+00;
  Matches  2883;  Conservative    3;  Mismatches 364;  Indels   4;  Gaps   2;

Db      1 GCTTGTCCTGAGAAGTGTCGCTGTGAAGGGACCACAGTAGACTGCTCCAATCAAAGACTC 60
          ||||| ||||| |||||||||||||||||| |||||||||| ||||| |||||||  |||
Qy   1501 gcttgccctgaaaagtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagctc 1560

Db     61 AACAAAATCCCTGACCATATTCCCCAGTACACAGCAGAGCTGCGTCTCAATAATAATGAA 120
          |||||||||| || || |||||||||||||| |||||| ||||||||||||||||||||
Qy   1561 aacaaaatcccggagcacattccccagtacactgcagagttgcgtctcaataataatgaa 1620

Db    121 TTCACAGTGTTAGAAGCCACGGGAATATTTAAGAAACTTCCTCAGTTACGTRAAATCAAC 180
          || || ||||| |||||||| ||||| ||||||||||||||||| ||||||:|||| |||
Qy   1621 tttaccgtgttggaagccacaggaatctttaagaaacttcctcaattacgtaaaataaac 1680

Db    181 TTTAGCAACAATAAGATCACGGATATCGAGGAGGGTGCATTTGAAGGTGCGTCTGGTGTG 240
          |||||||||||||||||| ||||| |||||||| |||||||||| || ||||||||
Qy   1681 tttagcaacaataagatcacagatattgaggagggagcatttgaaggagcatctggtgta 1740

Db    241 AATGAAATTCTTCTCACCAGTAACCGTTTGGAAAATGTTCAGCATAAGATGTTCAAAGGA 300
          |||||||| ||||| || ||||| |||||||||||||| ||||||||||||||||||| |||
Qy   1741 aatgaaatacttcttacgagtaatcgtttggaaaatgtgcagcataagatgttcaaggga 1800

Db    301 CTGGAGAGCCTCAAAACATTGATGCTGAGAAGTAATCGAATAAGCTGTGTTGGGAACGAC 360
           |||| |||||||||||| |||||| ||||||| |||||||||| |||||| ||||| |||
Qy   1801 ttggaaagcctcaaaactttgatgttgagaagcaatcgaataacctgtgtggggaatgac 1860

Db    361 AGTTTCATAGGACTCGGCTCTGTGCGTCTGCTCTCTTTATATGACAATCAAATTACCACA 420
          ||||||||||||||| | ||||||||| |||| ||||| ||||| |||||||||||| |||
Qy   1861 agtttcataggactcagttctgtgcgtttgctttctttgtatgataatcaaattactaca 1920

Db    421 GTGGCACCAGGAGCATTTGATYCTCTCCATTCATTATCCACTCTAAACCTCTTGGCCAAT 480
          || |||||||| |||||||||| |||||||||| ||||| ||||||||||||||||||||
Qy   1921 gttgcaccaggggcatttgatactctccattctttatctactctaaacctcttggccaat 1980

Db    481 CCTTTCAACTGTAACTGTCACCTGGCATGGCTGGGAGAATGGCTCAGAAGGAAAAGAATT 540
          ||||| |||||||||||  ||||||| ||| ||||||| ||||| |||| ||| ||||||
Qy   1981 cctttaactgtaactgctacctggcttggttgggagagtggctgagaaagaagagaatt 2040

Db    541 GTAACAGGAAATCCTCGATGCCAAAAACCCTACTTCCTGAAGGAAATCCCAATCCAGGAT 600
          || || ||||||||| |||| |||||||| |||||||||| ||||| || |||||||||
Qy   2041 gtcacgggaaatcctagatgtcaaaaaccatacttcctgaaagaaatacccatccaggat 2100

Db    601 GTAGCCATTCAGGACTTCACCTGTGATGATGGAAATGATGACAATAGTTGCTCTCCACTC 660
          || |||||||||||||||| |||||||| |||||||||||||||||||||||| |||||
```

```
Qy  2101 gtggccattcaggacttcacttgtgatgacggaaatgatgacaatagttgctccccactt 2160

Db   661 TCCCGTTGTCCTTCTGAATGTACCTGCTTGGATACAGKGGTACGATGTAGCAACAAGGGC 720

Qy  2161 tctcgctgtcctactgaatgtacttgcttggatacagtcgtccgatgtagcaacaagggt 2220

Db   721 TTGAAGGTCTTGCCTAAAGGTATTCCAAAAGATGTCACAGAGCTGTATCTGGATGGGAAC 780

Qy  2221 ttgaaggtcttgccgaaaggtattccaagagatgtcacagagttgtatctggatggaaac 2280

Db   781 CAGTTTACGCTGGTCCCGAAGGAACTCTCTAACTACAAACATTTAACACTTATAGACTTA 840

Qy  2281 caatttacactggttcccaaggaactctccaactacaaacatttaacacttatagactta 2340

Db   841 AGTAACAACCGAATAAGCACCCTTTCCAATCAAMGCTTCAGCAACATGACCCAGCTTCTC 900

Qy  2341 agtaacaacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcctc 2400

Db   901 ACCTTAATCCTCAGTTACAACCGTCTGAGATGTATCCCTCCACGAACCTTTGATGGATTG 960

Qy  2401 accttaattcttagttacaaccgtctgagatgtattcctcctcgcacctttgatggatta 2460

Db   961 AAGTCTCTTCGGTTACTGTCTTTACATGGAAATGACATTTCTGTTGTGCCTGAAGGTGCC 1020

Qy  2461 aagtctcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgct 2520

Db  1021 TTCAATGACTTGTCAGCCTTGTCACACTTAGCGATTGGAGCCAACCCTCTTTACTGTGAT 1080

Qy  2521 ttcaatgatctttctgcattatcacatctagcaattggagccaaccctctttactgtgat 2580

Db  1081 TGTAACATGCAGTGGTTATCCGACTGGGTGAAGTCGGAATATAAGGAACCTGGAATTGCA 1140

Qy  2581 tgtaacatgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgct 2640

Db  1141 CGCTGTGCCGGCCCTGGAGAAATGGCAGATAAATTATTACTCACTACTCCCTCCAAAAAA 1200

Qy  2641 cgttgtgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaaa 2700

Db  1201 TTTACATGTCAAGGTCCCATGGATATCACTATTCAAGCCAAGTGTAATCCCTGCTTATCA 1260

Qy  2701 tttacctgtcaaggtcctgtggatgtcaatattctagctaagtgtaaccccctgcctatca 2760

Db  1261 AATCCATGTAAAAATGATGGCACCTGTAACAATGACCCCGTTGATTTTTATCGATGTACC 1320

Qy  2761 aatccgtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcacc 2820

Db  1321 TGCCCATATGGATTCAAGGGTCAGGACTGTGATGTCCCCATTCATGCTTGTATCAGTAAT 1380

Qy  2821 tgtccatatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaac 2880

Db  1381 CCATGTAAACATGGAGGAACTTGTCACTTAAAGGAAGGAGAGAATGCTGGATTCTGGTGC 1440

Qy  2881 ccatgtaaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtgt 2940

Db  1441 ACTTGTGCTGATGGGTTTGAAGGAGAAAACTGTGAAGTCAATATTGATGATTGTGAAGAT 1500

Qy  2941 atttgtgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaagat 3000

Db  1501 AATGATTGTGAAAATAATTCTACATGCGTTGATGGAATTAACAACTACACATGTCTTTGC 1560

Qy  3001 aatgactgtgaaaataattctacatgtgtcgatggcattaataactacacatgcctttgc 3060

Db  1561 CCACCGGAATACACAGGTGAACTGTGTGAGGAAAAGCTGGACTTCTGTGCACAAGACTTG 1620

Qy  3061 ccacctgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacctg 3120

Db  1621 AATCCCTGCCAGCATGACTCCAAGTGCATCCTGACTCCAAAGGGATTCAAGTGTGACTGC 1680

Qy  3121 aacccctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactgc 3180

Db  1681 ACTCCAGGATACATTGGTGAGCACTGTGACATTGACTTTGATGACTGCCAAGATAACAAG 1740

Qy  3181 acaccagggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaag 3240

Db  1741 TGTAAAAACGGTGCTCACTGCACAGATGCCGTGAACGGATACACGTGCGTCTGTCCTGAA 1800

Qy  3241 tgtaaaaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccgaa 3300

Db  1801 GGCTACAGTGGCTTGTTCTGTGAGTTTTCTCCACCCATGGTCCTCCCTCGCACCAGCCCC 1860

Qy  3301 ggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagcccc 3360

Db  1861 TGTGATAATTTTGATTGCCAGAATGGAGCCCAGTGTATCATCAGGATAAATGAACCAATA 1920

Qy  3361 tgtgataattttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaata 3420

Db  1921 TGCCAGTGTTTGCCTGGCTACCTGGGAGAGAAGTGTGAGAAATTGGTCAGTGTGAATTTT 1980

Qy  3421 tgtcagtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaatttt 3480

Db  1981 GTAAACAAAGAGTCCTATCTTCAGATTCCTTCAGCCAAGGTTCGGCCTCAGACAAACATC 2040

Qy  3481 ataaacaaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacata 3540

Db  2041 ACACTTCAGATTGCCACAGATGAAGACAGCGGCATCCTCTTGTATAAAGGTGACAAAGAC 2100

Qy  3541 acacttcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaagac 3600

Db  2101 CACATTGCCGTGGAACTCTATCGAGGGCGAGTTCGAGCCAGCTATGACACCGGCTCTCAT 2160

Qy  3601 catatcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctcat 3660

Db  2161 CCGGCTTCTGCCATTTACAGTGTGGAGACAATCAATGATGGAAACTTCCACATTGTGGAG 2220

Qy  3661 ccagcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtggaa 3720

Db  2221 CTACTGACCCTGGATTCCAGTCTTTCCCTCTCTGTGGATGGAGGAAGCCCTAAAGTCATC 2280

Qy  3721 ctacttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcatc 3780

Db  2281 ACCAATTTGTCAAAACAATCTACTCTGAATTTCGACTCTCCACTCTATGTAGGAGGCATG 2340

Qy  3781 actaacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcatg 3840

Db  2341 CCTGGGAAAAATAACGTGGCATCCCTGCGCCAGGCCCCTGGGCAAAATGGCACCAGCTTC 2400

Qy  3841 ccagggaagagtaacgtggcatctctgcgccaggccctgggcagaacggaaccagcttc 3900

Db  2401 CATGGCTGTATCCGGAACCTTTACATTAACAGTGAGCTGCAGGACTTCCGGAAAATGCCT 2460

Qy  3901 cacggctgcatccggaacctttacatcaacagtgagctgcaggacttccagaaggtgccg 3960

Db  2461 ATGCAAACCGGAATTCTGCCTGGCTGTGAACCATGCCACAAGAAAGTATGTGCCCATGGC 2520

Qy  3961 atgcaaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatggc 4020

Db  2521 ATGTGCCAGCCCAGCAGCCAATCAGGCTTCACCTGTGAATGTGAGGAAGGGTGGATGGGG 2580

Qy  4021 acatgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatgggg 4080

Db  2581 CCCCTCTGTGACCAGAGAACCAATGATCCCTGCCTCGGAAACAAATGTGTGCATGGGACC 2640

Qy  4081 cccctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcacc 4140

Db  2641 TGCCTGCCCATCAATGCCTTCTCCTATAGTTGCAAGTGCCTGGAGGGCCATGGCGGTGTC 2700

Qy  4141 tgcttgcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgtc 4200

Db  2701 CTCTGTGATGAAGAAGAAGATCTCTTTAACCCCTGCCAGATGATCAAGTGCAAGCATGGG 2760

Qy  4201 ctctgtgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatggg 4260

Db  2761 AAGTGCAGGCTTTCTGGAGTGGGCCAGCCCTATTGTGAATGCAACAGTGGATTCACCGGG 2820

Qy  4261 aagtgcaggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacgggg 4320
```

```
Db   2821 GACAGCTGTGATAGAGAAATTTCTTGTCGAGGGGAACGGATAAGGGACTATTACCAGAAG 2880
          |||||||||||| ||||||| ||||||||||||||| ||||||| || |||||||| |||
Qy   4321 gacagctgtgatcgagaaatctcttgtcgaggggaaaggataagagattattaccaaaag 4380

Db   2881 CAGCAGGGTTACGCTGCCTGTCAAACAACTAAGAAAGTATCTCGCTTGGAATGCAGAGGC 2940
          |||||||| || ||||| || |||||||| ||||| || || || || || ||||||||
Qy   4381 cagcagggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagaggt 4440

Db   2941 GGGTGCGCTGGAGGCCAGTGCTGTGGACCTCTGAGAAGCAAGAGGCGGAAATACTCTTTC 3000
          ||||| || ||||| |||||||||||||| ||||| |||||| ||||||||||||||||
Qy   4441 gggtgtgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactctttc 4500

Db   3001 GAATGCACAGATGGCTCCTCATTTGTGGACGAGGTTGAGAAAGTGGTGAAGTGCGGCTGC 3060
          |||||||| || |||||||| ||||||||||||||||||||||||||||||||||||||
Qy   4501 gaatgcactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctgt 4560

Db   3061 GCGAGATGTGCCTCCTAAGCGCGTCTCTAGAAGCTTCTAGCTTCGGCGAAGGTTGTACAC 3120
           |||| ||||  |||||| | |    |  | |||| ||  ||| ||  |||||||||| ||
Qy   4561 acgaggtgtgtgtcctaaacacactcccggcagct-ctgtctttggaaaaggttgtatac 4619

Db   3121 TTCTTGACCATGTTGGACTAATTCATGCTTCATAATGGAAATATTTGAAATATATTGTAA 3180
          |||||||||||| ||||||||  |||||||||| ||||||||||||||||||||||||||
Qy   4620 ttcttgaccatgtgggactaatgaatgcttcatagtggaaatatttgaaatatattgtaa 4679

Db   3181 AATACAGAACAGACTTATTTTTATTATGATAATAAAGACTTGT---CTGCATTTGGAAAA 3237
          |||||||||||||||||||||||||||| |||||||||| |   ||||||||||||||
Qy   4680 aatacagaacagacttatttttattatgagaataaagacttttttttctgcatttggaaaa 4739

Db   3238 AAATAATAATAAAA 3251
          ||| || || ||||
Qy   4740 aaaaaaaaaaaaa 4753


RESULT    4
LOCUS       AB011531     5210 bp    mRNA           ROD       22-AUG-1998
DEFINITION  Rattus norvegicus mRNA for MEGF5, complete cds.
ACCESSION   AB011531
NID         g3449291
VERSION     AB011531.1  GI:3449291
KEYWORDS    MEGF5.
SOURCE      Rattus norvegicus (strain:Sprague-Dawley) adult brain cDNA to mRNA,
            clone_lib:pSPORT 1 clone:RG2635.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 5210)
  AUTHORS   Nakayama,M., Nakajima,D. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1998) to the DDBJ/EMBL/GenBank databases. Manabu
            Nakayama, Kazusa DNA Research Institute, Laboratory of DNA
            technology; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:nmanabu@kazusa.or.jp, Tel:+81-438-52-3915,
            Fax:+81-438-52-3914)
REFERENCE   2  (sites)
  AUTHORS   Nakayama,M., Nakajima,D., Nagase,T., Nomura,N., Seki,N. and
            Ohara,O.
  TITLE     Identification of high-molecular-weight proteins with multiple
            EGF-like motifs by motif-trap screening
  JOURNAL   Genomics 51 (1), 27-34 (1998)
  MEDLINE   98360089
FEATURES             Location/Qualifiers
     source          1. .5210
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="RG2635"
                     /clone_lib="pSPORT 1"
                     /dev_stage="adult"
                     /tissue_type="brain"
     gene            292. .4863
                     /gene="MEGF5"
     CDS             292. .4863
                     /gene="MEGF5"
                     /note="rat homologue of Drosophila slit protein; rat
                     slit2"
                     /codon_start=1
                     /product="MEGF5"
                     /protein_id="BAA32461.1"
                     /db_xref="PID:d1033424"
                     /db_xref="PID:g3449292"
                     /db_xref="GI:3449292"
                     /translation="MAPGRTGAGAAVRARLALALALASILSGPPAAACPTKCTCSAAS
                     VDCHGLGLRAVPRGIPRNAERLDLDRNNITRITKMDFTGLKNLRVLHLEDNQVSVIER
                     GAFQDLKQLERLRLNKNKLQVLPELLFQSTPKLTRLDLSENQIQGIPRKAFRGVTGVK
                     NLQLDNNHISCIEDGAFRALRDLEILTLNNNNISRILVTSFNHMPKIRTLRLHSNHLY
                     CDCHLAWLSDWLRQRRTIGQFTLCMAPVHLRGFSVADVQKKEYVCPGPHSEAPACNAN
                     SLSCPSACSCSNNIVDCRGKGLTEIPANLPEGIVEIRLEQNSIKSIPAGAFIQYKKLK
                     RIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIPKGLFDGLVSLQLLLLNANKIN
                     CLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHLAQNPFVCDCHLKWLAD
                     YLQDNPIETSGARCSSPRRLANKRISQIKSKKFRCSGSEDYRNRFSSECFMDLVCPEK
                     CRCEGTIVDCSNQKLSRIPSHLPEYTTDLRLNDNDIAVLEATGIFKKLPNLRKINLSN
                     NRIKEVREGAFDGAAGVQELMLTGNQLETMHGRMFRGLSGLKTLMLRSNLISCVNNDT
                     FAGLSSVRLLSLYDNRITTISPGAFTTLVSLSTINLLSNPFNCNCHMAWLGRWLRKRR
                     IVSGNPRCQKPFFLKEIPIQDVAIQDFTCEGNEENSCQLSPRCPEQCTCVETVVRCSN
                     RGLHTLPKGMPKDVTELYLEGNHLTAVPKELSTFRQLTLIDLSNNSISMLTNHTFSNM
                     SHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSLSHLALGI
                     NPLHCDCSLRWLSEWIKAGYKEPGIARCSSPESMADRLLLTTPTHRFQCKGPVDINIV
                     AKCNACLSSPCKNNGTCSQDPVEQYRCTCPYSYKGKDCTVPINTCVQNPCQHGGTCHL
                     SESHRDGFSCSCPLGFEGQRCEINPDDCEDNDCENSATCVDGINNYACVCPPNYTGEL
                     CDEVIDYCVPEMNLCQHEAKCISLDKGFRCECVPGYSGKLCETDNDDCVAHKCRHGAQ
                     CVDAVNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQQEPTCRCP
                     PGFAGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYKGDNDPL
                     ALELYQGHVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVMLNQTLNLVVDKGAPKSL
                     GKLQKQPAVGINSPLYLGGIPTSTGLSALRQGADRPLGGPHGCIHEVRINNELQDFKA
                     LPPQSLGVSPGCKSCTVCRHGLCRSVEKDSVVCECHPGWTGPLCDQEAQDPCLGHSCS
                     HGTCVATGNSYVCKCAEGYEGPLCDQKNDSANACSAFKCHHGQCHISDRGEPYCLCQP
                     GFSGNHCEQENPCLGEIVREAIRRQKDYASCATASKVPIMVCRGGCGSQCCQPIRSKR
                     RKYVFQCTDGSSFVEEVERHLECGCRECS"
     polyA_site      5210
                     /note="17 a nucleotides"
BASE COUNT     1210 a   1583 c   1386 g   1031 t
ORIGIN


  Query Match          26.3%;  Score 1249;  DB 32;  Length 5210;
  Best Local Similarity  65.1%;  Pred. No. 0.00e+00;
  Matches  2918;  Conservative    0;  Mismatches 1543;  Indels  21;  Gaps  17;

Db    388 GCCTGCCCCACCAAGTGTACCTGCTCCGCCGCTAGCGTGGACTGCCACGGGCTGGGCCTG 447
          || |||||  |  ||||  | |||||  | |   |  |||||||| ||||||||||  |||
Qy     79 gcgtgcccggcgcagtgctcttgctcgggcagcacagtggactgtcacgggctggcgctg 138

Db    448 CGCGCCGTTCCCCCGGGGCATCCCCCCGCAACGCTGAGCCGCCTTGACCTGGACAGAAATAAC 507
          |||  ||| ||| ||   |||||||||||| | ||| | || ||  |  |  ||||||||
Qy    139 cgcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatggaaataac 198

Db    508 ATCACCAGGATCACCAAAATGGACTTCACCGGCTTGAAGAATCTCCGAGTCTTGCATCTG 567
          ||||| || || || || |  || ||  | ||  | |   ||||  ||||  | || ||
Qy    199 atcacaagaattacgaagacagattttgctggtcttagacatctaagagttcttcagctt 258

Db    568 GAAGACAATCAGGTCAGCGTCATCGAGAGAGGCGCCTTCCAGGATCTGAAGCAGCTGGAG 627
            || ||| || | |||  ||| || ||||| || ||||||||||| ||  | || |||
Qy    259 atggagaataagattagcaccattgaaagaggagcattccaggatcttaaagaactagag 318

Db    628 CGATTACGTCTGAACAAGAACAAGCTCCAGGTCCTTCCAGAATTACTTTTCCAGAGCACA 687
          || | ||| | ||||  ||  | || ||| |  |||| || || || || |   | ||
Qy    319 agactgcgtttaaacagaaatcaccttcagctgtttcctgagttgctgtttcttgggact 378

Db    688 CCGAAGCTCACCAGACTAGATCTGAGCGAAAACCAGATCCAGGGCATCCCGAGGAAGGCG 747
          |||||||   ||| || ||||| || |||||||| || ||||  ||||| ||||| ||
Qy    379 gcgaagctatacaggcttgatctcagtgaaaaccaaattcaggcaatcccaaggaaagct 438

Db    748 TTCAGGGGCGTCACGGGCGTGAAGAACCTGCAACTGGACAATAACCACATCAGCTGCATT 807
          ||| | || |     | | | || ||  ||||||||||  | ||||| ||||||| |||
Qy    439 ttccgtggggcagttgacataaaaaatttgcaactggattacaaccagatcagctgtatt 498

Db    808 GAAGATGGAGCCTTCCGAGCGCTGCGCGATTTGGAGATCCTCACCCTTAACAACAACAAC 867
```

```
Qy   499 gaagatggggcattcagggctctccgggacctggaagtgctcactctcaacaataacaac 558

Db   868 ATCAGCCGTATCCTGGTCACCAGCTTCAACCACATGCCGAAGATCCGGACCCTGCGTCTC 927

Qy   559 attactagactttctgtggcaagtttcaaccatatgcctaaacttaggacttttcgactg 618

Db   928 CACTCCAACCACCTGTACTGTGATTGTCACTTGGCCTGGCTCTCAGACTGGCTGCGACAG 987

Qy   619 cattcaaacaacctgtattgtgactgccacctggcctggctctccgactggcttcgcaaa 678

Db   988 CGGCGGACCATTGGCCAGTTCACCCTCTGCATGGCGCCCGTGCACCTGAGAGGCTTCAGT 1047

Qy   679 aggcctcgggttggtctgtacactcagtgtatgggcccctcccacctgagaggccataat 738

Db  1048 GTAGCAGACGTGCAGAAGA-AGGA---GTATGT-GTGTCCAGGTCCCCACTCAG-AGGC- 1100

Qy   739 gtagccgaggttcaaaaacgagaatttgtctgcagtgatgaggaagaaggtcaccagtca 798

Db  1101 TCCA-GCCTGCAATGCCAACTCCCTCTC-CTGCCCTTCTGCCTGCTCATGCAGTAATAAC 1158

Qy   799 tttatggctccttcttgtagtgttttgcactgccctgccgcctgtacctgtagcaacaat 858

Db  1159 ATTGTAGACTGCCGTGGGAAGGGACTGACTGAGATCCCTGCCAACCTGCCGGAGGGCATC 1218

Qy   859 atcgtagactgtcgtgggaaaggtctcactgagatccccacaaatcttccagagaccatc 918

Db  1219 GTGGAAATACGCCTAGAACAGAACTCCATCAAATCCATTCCTGCAGGAGCTTTCATCCAG 1278

Qy   919 acagaaatacgtttggaacagaacacaatcaaagtcatccctcctggagctttctcacca 978

Db  1279 TACAAAAAACTGAAGCGAATAGACATCAGCAAGAATCAGATATCGGACATTGCTCCAGAT 1338

Qy   979 tataaaaagcttagacgaattgacctgagcaataatcagatctctgaacttgcaccagat 1038

Db  1339 GCCTTCCAGGGCCTGAAATCACTCACGTCGCTGGTGCTCTATGGGAACAAGATCACAGAG 1398

Qy  1039 gctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcacagaa 1098

Db  1399 ATTCCCAAAGGACTGTTTGACGGGCTGGTGTCCCTGCAGCTGCTCCTCCTCAATGCCAAC 1458

Qy  1099 ctccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgccaac 1158

Db  1459 AAGATCAACTGCCTGCGGGTAAACACCTTCCAGGACCTACAGAACCTCAATCTGCTCTCT 1518

Qy  1159 aagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctctcc 1218

Db  1519 CTGTATGACAACAAGTTGCAGACTATCAGCAAAGGGCTCTTTGCTCCGCTGCAGTCCATC 1578

Qy  1219 ctatatgacaacaagcttcagaccatcgccaaggggaccttttcacctcttcgggccatt 1278

Db  1579 CAGACCCTCCACTTAGCTCAAAACCCGTTTGTTTGCGACTGCCACTTGAAGTGGTTGGCC 1638

Qy  1279 caaactatgcatttggcccagaaccccttttatttgtgactgccatctcaagtggctagcg 1338

Db  1639 GACTACCTCCAGGACAACCCCATTGAGACGAGCGGGGCCCGCTGCAGCAGCCCACGCCGG 1698

Qy  1339 gattatctccataccaacccgattgagaccagtggtgcccgttgcaccagcccccgccgc 1398

Db  1699 CTGGCCAACAAGCGCATCAGCCAGATCAAAAGCAAGAAGTTCCGCTGCTCAGGCTCGGAG 1758

Qy  1399 ctggcaaacaaaagaattggacagatcaaaagcaagaaattccgttgttcaggtacagaa 1458

Db  1759 GATTATCGCAACAGATTCAGCAGCGAGTGCTTCATGGACCTAGTGTGCCCCGAGAAGTGC 1818

Qy  1459 gattatcgatcaaaattaagtggagactgctttgcggatctggcttgccctgaaaagtgt 1518

Db  1819 CGTTGTGAGGGCACCATTGTGGACTGCTCCAACCAGAAGCTCTCCCGCATCCCGAGCCAC 1878

Qy  1519 cgctgtgaaggaaccacagtagattgctctaatcaaaagctcaacaaaatcccggagcac 1578

Db  1879 CTCCCTGAATATACCACTGACCTGCGACTGAACGACAATGACATCGCTGTGCTGGAGGCC 1938

Qy  1579 attccccagtacactgcagagttgcgtctcaataataatgaatttaccgtgttggaagcc 1638

Db  1939 ACTGGGATCTTCAAGAAGTTGCCCAACCTGAGGAAAATAAACTTGAGCAATAATAGGATC 1998

Qy  1639 acaggaatctttaagaaacttcctcaattacgtaaaataaacttagcaacaataagatc 1698

Db  1999 AAGGAGGTGCGGGAGGGTGCGTTTGATGGAGCGGCAGGCGTGCAGGAACTGATGCTGACG 2058

Qy  1699 acagatattgaggagggagcatttgaaggagcatctggtgtaaatgaaatacttcttacg 1758

Db  2059 GGGAACCAGCTGGAGACCATGCACGGACGCATGTTCCGGGGCCTCAGCGGCCTGAAGACA 2118

Qy  1759 agtaatcgtttggaaaatgtgcagcataagatgttcaagggattggaaagcctcaaaact 1818

Db  2119 CTGATGTTAAGGAGCAACCTGATCAGCTGCGTAAACAATGACACCTTCGCTGGCCTAAGC 2178

Qy  1819 ttgatgttgagaagcaatcgaataacctgtgtggggaatgacagtttcataggactcagt 1878

Db  2179 TCTGTGAGACTGCTGTCCCTCTATGACAATCGGATCACCACCATCAGCCCTGGAGCCTTC 2238

Qy  1879 tctgtgcgtttgctttctttgtatgataatcaaattactacagttgcaccaggggcattt 1938

Db  2239 ACCACGCTCGTCTCCCTGTCCACCATAAACCTCTTGTCTAACCCTTTCAACTGCAACTGT 2298

Qy  1939 gatactctccattctttatctactctaaacctcttggccaatccttttaactgtaactgc 1998

Db  2299 CACATGGCCTGGCTCGGCCGGTGGCTGAGGAAACGCCGCATCGTAAGCGGGAACCCCAGA 2358

Qy  1999 tacctggcttggttgggagagtggctgagaaagaagagaattgtcacgggaaatcctaga 2058

Db  2359 TGTCAGAAGCCCTTTTTCCTCAAGGAGATTCCTATCCAAGATGTGGCCATCCAGGACTTC 2418

Qy  2059 tgtcaaaaaccatacttcctgaaagaaatacccatccaggatgtggccattcaggacttc 2118

Db  2419 ACCTGTGA--A-GGCAATGAAGAGAACAGCTGTCAGCTGAGTCCACGCTGCCCCGAGCAG 2475

Qy  2119 acttgtgatgacggaaatgatgacaatagttgctccccactttctcgctgtcctactgaa 2178

Db  2476 TGTACCTGTGTGGAGACAGTGGTGCGATGCAGCAACAGGGGTCTCCACACCCTCCCCAAG 2535

Qy  2179 tgtacttgcttggatacagtcgtccgatgtagcaacaagggtttgaaggtcttgccgaaa 2238

Db  2536 GGCATGCCAAAGGACGTGACTGAACTGTACCTGGAAGGAAATCATTTAACGGCGGTGCCC 2595

Qy  2239 ggtattccaagagatgtcacagagttgtatctggatggaaaccaatttacactggttccc 2298

Db  2596 AAAGAATTGTCCACCTTCCGACAGCTGACACTAATTGACCTGAGCAACAACAGCATCAGT 2655

Qy  2299 aaggaactctccaactacaaacatttaacacttatagacttaagtaacaacagaataagc 2358

Db  2656 ATGCTGACCAATCACACCTTCAGCAACATGTCCCACCTCTCCACACTGATCCTGAGCTAC 2715

Qy  2359 acgctttctaatcagagcttcagcaacatgacccagctcctcaccttaattcttagttac 2418

Db  2716 AACCGGCTGAGATGCATCCCGGTCCATGCCTTCAACGGGCTAAGGTCACTCCGAGTGCTA 2775

Qy  2419 aaccgtctgagatgtattcctcctcgcaccttttgatggattaaagtctcttcgattactt 2478

Db  2776 ACCCTCCATGGCAATGACATTTCCAGTGTTCCTGAAGGCTCCTTCAATGATCTGACGTCC 2835

Qy  2479 tctctacatggaaatgacatttctgttgtgcctgaaggtgctttcaatgatctttctgca 2538

Db  2836 CTCTCCCACCTGGCCCTGGGAATCAACCCTCTCCACTGTGACTGCAGTCTGCGTTGGTTA 2895

Qy  2539 ttatcacatctagcaattggagccaaccctctttactgtgattgtaacatgcagtggtta 2598

Db  2896 TCAGAGTGGATAAAGGCTGGGTACAAGGAGCCTGGCATTGCCAGATGCAGTAGCCCCGAG 2955

Qy  2599 tccgactgggtgaagtcggaatataaggagcctggaattgctcgttgtgctggtcctgga 2658

Db  2956 TCCATGGCAGACAGACTCCTGCTCACCACCCCCACCCACAGGTTCCAGTGCAAAGGGCCA 3015

Qy  2659 gaaatggcagataaacttttactcacaactcccctccaaaaaatttacctgtcaaggtcct 2718
```

```
Db  3016 GTGGACATTAACATCGTGGCCAAGTGCAACGCCTGCCTCTCTAGCCCATGCAAGAACAAC 3075
Qy  2719 gtggatgtcaatattctagctaagtgtaacccctgcctatcaaatccgtgtaaaaatgat 2778

Db  3076 GGCACTTGCAGCCAGGATCCGGTGGAGCAGTACCGCTGTACCTGCCCTTACAGCTACAAG 3135
Qy  2779 ggcacatgtaatagtgatccagttgacttttaccgatgcacctgtccatatggtttcaag 2838

Db  3136 GGCAAGGACTGCACCGTGCCCATCAACACCTGCGTCCAGAACCCCTGTCAGCACGGAGGC 3195
Qy  2839 gggcaggactgtgatgtcccaattcatgcctgcatcagtaacccatgtaaacatggagga 2898

Db  3196 ACCTGTCACCTCAGTGAGAGCCACAGAGATGGGTTCAGCTGCTCCTGCCCCCTGGGCTTT 3255
Qy  2899 acttgccacttaaaggaaggagaagaagatggattctggtgtatttgtgctgatggattt 2958

Db  3256 GAGGGACAGCGGTGTGAGATCAACCCAGATGACTGTGAGGATAACGACTGTGAAAACAGC 3315
Qy  2959 gaaggagaaaattgtgaagtcaacgttgatgattgtgaagataatgactgtgaaaataat 3018

Db  3316 GCCACCTGTGTGGACGGGATCAACAACTACGCATGCGTGTGCCCGCCGAACTACACAGGG 3375
Qy  3019 tctacatgtgtcgatggcattaataactacacatgcctttgcccacctgagtatacaggt 3078

Db  3376 GAGCTGTGTGATGAGGTGATTGACTACTGCGTGCCCGAGATGAACCTCTGTCAGCACGAG 3435
Qy  3079 gagttgtgtgaggagaagctggacttctgtgcccaggacctgaacccctgccagcacgat 3138

Db  3436 GCCAAGTGCATCTCCCTGGACAAAGGATTCAGGTGCGAATGTGTCCCTGGCTACAGTGGA 3495
Qy  3139 tcaaagtgcatcctaactccaaagggattcaaatgtgactgcacaccagggtacgtaggt 3198

Db  3496 AAGCTGTGCGAGACAGACAACGATGACTGTGTGGCTCACAAGTGTCGCCACGGAGCCCAG 3555
Qy  3199 gaacactgcgacatcgattttgacgactgccaagacaacaagtgtaaaaacggagcccac 3258

Db  3556 TGTGTGGACGCGGTCAATGGCTACACGTGCATCTGCCCCCAGGGCTTCAGCGGGCTCTTC 3615
Qy  3259 tgcacagatgcagtgaacggctatacgtgcatatgccccgaaggttacagtggcttgttc 3318

Db  3616 TGTGAGCACCCCCCACCCATGGTTCTGCTACAGACCAGCCCCTGTGACCAGTACGAGTGC 3675
Qy  3319 tgtgagttttctccacccatggtcctccctcgtaccagcccctgtgataattttgattgt 3378

Db  3676 CAGAATGGGGCGCAGTGCATCGTGGTACAGCAGGAGCCCACTTGCCGCTGTCCCCCAGGC 3735
Qy  3379 cagaatggagctcagtgtatcgtcagaataaatgagccaatatgtcagtgtttgcctggc 3438

Db  3736 TTCGCTGGGCCCAGGTGTGAGAAGCTCATCACTGTGAATTTCGTGGGCAAGGACTCCTAT 3795
Qy  3439 tatcagggagaaaagtgtgaaaaattggttagtgtgaattttataaacaaagagtcttat 3498

Db  3796 GTGGAACTGGCCTCTGCCAAGGTCCGGCCCCAGGCCAACATCTCCCTGCAGGTGGCCACT 3855
Qy  3499 cttcagattccttcagccaaggttcggcctcagacgaacataacacttcagattgccaca 3558

Db  3856 GACAAGGACAACGGCATCCTTCTTTACAAGGGAGACAACGACCCCCTGGCACTGGAGCTG 3915
Qy  3559 gatgaagacagcggaatcctcctgtataagggtgacaaagaccatatcgcggtagaactc 3618

Db  3916 TACCAGGGTCACGTGAGGCTGGTGTATGACAGCCTGAGCTCCCCTCCGACCACGGTGTAC 3975
Qy  3619 tatcgggggcgtgttcgtgccagctatgacaccggctctcatccagcttctgccatttac 3678

Db  3976 AGTGTGGAGACGGTGAATGATGGCCAGTTTCACAGTGTGGAGCTGGTGATGCTAAACCAG 4035
Qy  3679 agtgtggagacaatcaatgatggaaacttccacattgtggaactacttgccttggatcag 3738

Db  4036 ACCCTGAACCTGGTGGTAGACAAAGGAGCCCCCAAGAGCCTGGGGAAGCTCCAGAAGCAG 4095
Qy  3739 agtctctctttgtccgtggatggtgggaaccccaaaatcatcactaacttgtcaaagcag 3798

Db  4096 CCAGCAGTGGGCATCAACAGTCCCCTCTATCTTGGAGGCATCCCCACGTCTACAGGCCTC 4155
Qy  3799 tccactctgaattttgactctccactctatgtaggaggcatgccagggaagagtaacgtg 3858

Db  4156 TCAGCCTTACGCCAGGGTGCAGACAGGCCGCTGGGCGGCTTCCACGGCTGTATACACGAA 4215
Qy  3859 gcatctctgcgccaggcccctgggcagaacggaaccagcttccacggctgcatccggaac 3918

Db  4216 GTGCGCATCAACAACGAGTTGCAGGATTTCAAAGCCCTCCCACCCCAGTCCCTGGGGGTC 4275
Qy  3919 ctttacatcaacagtgagctgcaggacttccagaaggtgccgatgcaaacaggcattttg 3978

Db  4276 TCTCCCGGCT-GCAAATCCTGCACT--GTGTGTCGTCACGGCCTGTGTCGTTCCGTGGAG 4332
Qy  3979 cctggctgtgagccatgccacaagaaggtgtgtgcccatggcacatgccagcccagcagc 4038

Db  4333 AAGGACAGCGTAGTGTGTGAGTGCCATCCGGGATGGACTGGTCCACTGTGTGATCAGGAA 4392
Qy  4039 caggcaggcttcacctgcgagtgccaggaaggatggatggggccccctctgtgaccaacgg 4098

Db  4393 GCCCAGGACCCCTGCCTTGGTCACAGCTGCAGCCATGGGACATGCGTGGCGA-CTG-GCA 4450
Qy  4099 accaatgacccttgccttggaaataaatgcgtacatggcacctgcttgcccatcaatgcg 4158

Db  4451 A-CTCATATGTGTGCAAGTGTGCCGAGGGCTATGAGGGACCTTTGTGTGACCAGAAGAAT 4509
Qy  4159 ttctcctacagctgtaagtgcttggagggccatggaggtgtcctctgtgatgaagaggag 4218

Db  4510 GACTCTGCCAATGCCTGCTCAGCCTTCAAGTGCCACCACGGGCAGTGCCACATCTCAGAT 4569
Qy  4219 gatctgtttaacccatgccaggcgatcaagtgcaagcatgggaagtgcaggctttcaggt 4278

Db  4570 CGAGGGGAGCCCTATTGCCTGTGCCAGCCTGGCTTCAGTGGCAATCACTGTGAGCAAGAG 4629
Qy  4279 ctggggcagccctactgtgaatgcagcagtggatacacgggggacagctgtgatcgagaa 4338

Db  4630 AATCCATGTTTGGGAGAGATAGTCCGTGAAGCCATCCGCCGCCAGAAAGATTATGCCTCC 4689
Qy  4339 atctcttgtcgaggggaaaggataagagattattaccaaaagcagcagggctatgctgct 4398

Db  4690 TGTGCCACAGCATCCAAGGTGCCCATCATGGTATGCCGCGG-GGGC-TGCGGGAGC-CAA 4746
Qy  4399 tgccaaacaaccaagaaggtgtcccgattagagtgcagaggtgggtgtgcaggagggcag 4458

Db  4747 TGCTGCCAGCCGATTCGGAGCAAGCGGCGGAAATACGTCTTCCAGTGCACTGACGGCTCC 4806
Qy  4459 tgctgtggaccgctgaggagcaagcggcggaaatactctttcgaatgcactgacggctcc 4518

Db  4807 TCGTTCGTGGAAGAGGTGGAGAGACACTTGGAATGTGGCTGT 4848
Qy  4519 tcctttgtggacgaggttgagaaagtggtgaagtgcggctgt 4560
```

```
RESULT    5
LOCUS       AB017167     5094 bp    mRNA            PRI       06-FEB-1999
DEFINITION  Homo sapiens mRNA for Slit-1 protein, complete cds.
ACCESSION   AB017167
NID         g4049584
VERSION     AB017167.1  GI:4049584
KEYWORDS    slit-1; Slit-1 protein.
SOURCE      Homo sapiens fetal tissue_lib:brain cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5094)
  AUTHORS   Itoh,A. and Sakano,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1998) to the DDBJ/EMBL/GenBank databases. Akira
            Itoh, Asahi Chemical Industry co.,ltd., Life Science Fundamental
            Research Laboratory; 2-1, Samejima, Fuji, Shizuoka 416-8501, Japan
            (E-mail:a8611483@ut.asahi-kasei.co.jp, Tel:+81-545-62-3231,
            Fax:+81-545-62-3249)
REFERENCE   2  (sites)
```

```
  AUTHORS   Itoh,A., Miyabayashi,T., Ohno,M. and Sakano,S.
  TITLE     Cloning and expressions of three mammalian homologues of Drosophila
            slit suggest possible roles for Slit in the formation and
            maintenance of the nervous system
  JOURNAL   Brain Res. Mol. Brain Res. 62 (2), 175-186 (1998)
  MEDLINE   99033071
FEATURES             Location/Qualifiers
     source          1. .5094
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="fetal"
                     /tissue_lib="brain"
     gene            233. .4837
                     /gene="slit-1"
     CDS             233. .4837
                     /gene="slit-1"
                     /note="human homologue of Drosophila slit gene"
                     /codon_start=1
                     /product="Slit-1 protein"
                     /protein_id="BAA35184.1"
                     /db_xref="PID:d1036170"
                     /db_xref="PID:g4049585"
                     /db_xref="GI:4049585"
                     /translation="MALTPGWGSSAGPVRPELWLLLWAAAWRLGASACPALCTCTGTT
                     VDCHGTGLQAIPKNIPRNTERLELNGNNITRIHKNDFAGLKQLRVLQLMENQIGAVER
                     GAFDDMKELERLRLNRNQLHMLPELLFQNNQALSRLDLSENAIQAIPRKAFRGATDLK
                     NLRLDKNQISCIEEGAFRALRGLEVLTLNNNNITTIPVSSFNHMPKLRTFRLHSNHLF
                     CDCHLAWLSQWLRQRPTIGLFTQCSGPASLRGLNVAEVQKSEFSCSGQGEAGRVPTCT
                     LSSGSCPAMCTCSNGIVDCRGKGLTAIPANLPETMTEIRLELNGIKSIPPGAFSPYRK
                     LRRIDLSNNQIAEIAPDAFQGLRSLNSLVLYGNKITDLPRGVFGGLYTLQLLLLNANK
                     INCIRPDAFQDLQNLSLLSLYDNKIQSLAKGTFTSLRAIQTLHLAQNPFICDCNLKWL
                     ADFLRTNPIETSGARCASPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYQLNSECN
                     SDVVCPHKCRCEANVVECSSLKLTKIPERIPQSTAELRLNNNEISILEATGMFKKLTH
                     LKKINLSNNKVSEIEDGAFEGAASVSELHLTANQLESIRSGMFRGLDGLRTLMLRNNR
                     ISCIHNDSFTGLRNVRLLSLYDNQITTVSPGAFDTLQSLSTLNLLANPFNCNCQLAWL
                     GGWLRKRKIVTGNPRCQNPDFLRQIPLQDVAFPDFRCEEGQEEGGCLPRPQCPQECAC
                     LDTVVRCSNKHLRALPKGIPKNVTELYLDGNQFTLVPGQLSTFKYLQLVDLSNNKISS
                     LSNSSFTNMSQLTTLILSYNALQCIPPLAFQGLRSLRLLSLHGNDISTLQEGIFADVT
                     SLSHLAIGANPLYCDCHLRWLSSWVKTGYKEPGIARCAGPQDMEGKLLLTTPAKKFEC
                     QGPPTLAVQAKCDLCLSSPCQNQGTCHNDPLEVYRCACPSGYKGRDCEVSLNSCSSGP
                     CENGGTCHAQEGEDAPFTCSCPTGFEGPTCGVNTDDCVDHACANGGVCVDGVGNYTCQ
                     CPLQYEGKACEQLVDLCSPDLNPCQHEAQCVGTPDGPRCECMPGYAGDNCSENQDDCR
                     DHRCQNGAQCMDEVNSYSCLCAEGYSGQLCEIPPHLPAPKSPCEGTECQNGANCVDQG
                     NRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANITLQVSTAEDNGILL
                     YNGDNDHIAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVAFDQMVNLSI
                     DGGSPMTMDNFGKHYTLNSEAPLYVGGMPVDVNSAAFRLWQILNGTGFHGCIRNLYIN
                     NELQDFTKTQMKPGVVPGCEPCRKLYLCHGICQPNATPGPMCHCEAGWVGLHCDQPAD
                     GPCHGHKCVHGQCVPLDALSYSCQCQDGYSGALCNQAGALAEPCRGLQCLHGHCQASG
                     TKGAHCVCDPGFSGELCEQESECRGDPVRDFHQVQRGYAICQTTRPLSWVECRGSCPG
                     QGCCQGLRLKRRKFTFECSDGTSFAEEVEKPTKCGCALCA"
BASE COUNT     1037 a   1616 c   1508 g    933 t
ORIGIN
```

```
  Query Match           25.9%;  Score 1231;  DB 29;  Length 5094;
  Best Local Similarity  65.0%;  Pred. No. 0.00e+00;
  Matches  2926;  Conservative    0;  Mismatches 1551;  Indels  24;  Gaps  17;

Db    328 GGCGTGCCCCGCCCTCTGCACCTGCACCGGAACCACGGTGGACTGCCACGGCACGGGGCT 387
            ||||||||| || |  ||| | ||| | || | ||| |||||||| |||||   || |||
Qy     78 ggcgtgcccggcgcagtgctcttgctcgggcagcacagtggactgtcacgggctggcgct 137

Db    388 GCAGGCCATTCCCAAGAATATACCTCGGAACACCGAGCGCCTGGAACTCAATGGCAACAA 447
            ||    | | |||| |||||| || || |||||||||| | |||||  | |||||| || ||
Qy    138 gcgcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatggaaataa 197

Db    448 CATCACTCGGATCCATAAGAATGACTTTGCCGGGGCTCAAGCAGCTGCGGGTGCTGCAGCT 507
            ||||||  | ||    ||||  || ||||| || || |  || ||  | || || |||||
Qy    198 catcacaagaattacgaagacagattttgctggtcttagacatctaagagttcttcagct 257

Db    508 GATGGAGAACCAGATTGGAGCAGTGGAACGTGGTGCTTTTGATGACATGAAGGAGCTGGA 567
            ||||||||  |||||  |  |  | ||| | || || ||  | ||  | || || || ||
Qy    258 tatggagaataagattagcaccattgaaagaggagcattccaggatcttaaagaactaga 317
```

```
Db    568 GCGGCTGCGACTGAACCGAAACCAGCTGCACATGTTACCGGAACTGCTGTTCCAGAACAA 627
            | | |||||  | ||| |||| || || ||  |||| || ||  |||||||  |     |
Qy    318 gagactgcgtttaaacagaaatcaccttcagctgtttcctgagttgctgtttcttgggac 377

Db    628 CCAGGCTTTGTCAAGACTGGACTTGAGTGAGAACGCCATCCAGGCCATCCCCAGGAAAGC 687
              |   | |  || || ||  | ||||| |||   || ||||| ||||| ||||||||
Qy    378 tgcgaagctatacaggcttgatctcagtgaaaaccaaattcaggcaatcccaaggaaagc 437

Db    688 TTTTCGGGGAGCTACGGACCTTAAAAATTTACGGCTGGACAAGAACCAGATCAGCTGCAT 747
            ||| || || ||    ||| | |||||||| |  |||||  | ||||||||||||| ||
Qy    438 tttccgtggggcagttgacataaaaaatttgcaactggattacaaccagatcagctgtat 497

Db    748 TGAGGAAGGGGCCTTCCGTGCTCTGCGGGGGCTGGAGGTGCTGACCCTGAACAACAACAA 807
            ||| || ||||| ||| | |||||  ||||  |||||  ||||| || || |||||  |||||
Qy    498 tgaagatggggcattcagggctctccgggacctggaagtgctcactctcaacaataacaa 557

Db    808 TATCACCACCATCCCCGTGTCCAGCTTCAACCATATGCCCAAGCTACGGACCTTCCGCCT 867
            || || |   |  | ||| | || |||||||||||||| || ||  |||| || || ||
Qy    558 cattactagactttctgtggcaagtttcaaccatatgcctaaacttaggacttttcgact 617

Db    868 GCACTCCAACCACCTGTTTTGCGACTGCCACCTGGCCTGGCTCTCGCAGTGGCTGAGGCA 927
            ||| || ||| |||||| || |||||||||||||||||||||||  | |||||  |  |
Qy    618 gcattcaaacaacctgtattgtgactgccacctggcctggctctccgactggcttcgcaa 677

Db    928 GCGGCCAACCATCGGGCTCTTCACCCAGTGCTCGGGCCCAGCCAGCCTGCGTGGCCTCAA 987
              ||||     | || || | ||| |||||   ||||||  ||  |||| | ||||  ||
Qy    678 aaggcctcgggttggtctgtacactcagtgtatgggcccctcccacctgagaggccataa 737

Db    988 TGTGGCAGAGGTCCAGAAGAGTGAGTTCAGCTGCTCAGGCCAGGGAGAAGCGGGGC-G-C 1045
            ||| || |||||  || ||  | || ||   ||||   |   ||| |||||      | | |
Qy    738 tgtagccgaggttcaaaaacgagaatttgtctgcagtgatgaggaagaaggtcaccagtc 797

Db   1046 GTGCCCACCTGCACCCTGTCCTCCGGCTC-CTGCCCGGCCATGTGCACCTGCAGCAATGG 1104
            |      || |  | |||  |      | ||||||  |||   || |||||  |||||
Qy    798 atttatggctccttcttgtagtgttttgcactgccctgccgcctgtacctgtagcaacaa 857

Db   1105 CATCGTGGACTGTCGTGGAAAAGGCCTCACTGCCATCCCGGCCAACCTGCCCGAGACCAT 1164
            |||||  |||||||||||  |||||  |||||||  |||||  | || || || ||||||||
Qy    858 tatcgtagactgtcgtgggaaaggtctcactgagatccccacaaatcttccagagaccat 917

Db   1165 GACGGAGATACGCCTGGAGCTGAACGGCATCAAGTCCATCCCTCCTGGAGCCTTCTCACC 1224
            || || |||||  |||| | ||||  |||||  |||||||||||||||| ||||||||
Qy    918 cacagaaatacgtttggaacagaacacaatcaaagtcatccctcctggagctttctcacc 977

Db   1225 CTACAGAAAGCTACGGAGGATAGACCTGAGCAACAATCAGATCGCTGAGATTGCACCCGA 1284
            || | ||||||  |  | || ||||||||||| |||||||| ||||  ||||||| ||
Qy    978 atataaaaagcttagacgaattgacctgagcaataatcagatctctgaacttgcaccaga 1037

Db   1285 CGCCTTCCAGGGCCTCCGCTCCCTGAACTCGCTGGTCCTCTATGGAAACAAGATCACAGA 1344
            || ||||| || || ||||| ||||| || || |||||||||||||| || ||||||||
Qy   1038 tgctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcacaga 1097

Db   1345 CCTCCCCCGTGGTGTGTTTGGAGGCCTATACACCCTACAGCTCCTGCTCCTGAATGCCAA 1404
            ||||||    || | |||| ||| || |   || ||||||||||  |  ||||||||||
Qy   1098 actccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgccaa 1157

Db   1405 CAAGATCAACTGCATCCGGCCCGATGCCTTCCAGGACCTGCAGAACCTCTCACTGCTCTC 1464
            ||||||  |||||| | |||   |||| || |||| || || ||| |    || ||||||
Qy   1158 caagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctctc 1217

Db   1465 CCTGTATGACAACAAGATCCAGAGCCTCGCCAAGGGCACTTTCACCTCCCTGCGGGCCAT 1524
            ||| ||||||||||||  | |||| | ||||||||||  || ||   |  | || ||||||||
Qy   1218 cctatatgacaacaagcttcagaccatcgccaaggggaccttttcacctcttcgggccat 1277

Db   1525 CCAGACTCTGCACCTGGCCCAGAACCCTTTCATTTGCGACTGTAACCTCAAGTGGCTGGC 1584
            || ||| ||||  |||||||||||||| || |||||  |||||  | ||||||||||| ||
Qy   1278 tcaaactatgcatttggcccagaaccccttttatttgtgactgccatctcaagtggctagc 1337

Db   1585 AGACTTCCTGCGCACCAATCCCATCGAGACGAGTGGTGCCCGCTGTGCCAGTCCCCGGCG 1644
            || |  || |  ||||| || || ||||| ||||||||||| ||  |||| |||||  ||
Qy   1338 ggattatctccataccaacccgattgagaccagtggtgcccgttgcaccagcccccgccg 1397

Db   1645 CCTCGCCAACAAGCGCATCGGGCAGATCAAGAGCAAGAAGTTCCGGTGCTCAGCCAAAGA 1704
```

```
Qy  1398 cctggcaaacaaaagaattggacagatcaaaagcaagaaattccgttgttcaggtacaga 1457

Db  1705 GCAGTACTTCATTCCAGGCACGGAGGATTACCAGCTGAACAGCGAGTGCAACAGCGACGT 1764

Qy  1458 --ag-a-tt-at-cgat-caa--a--attaagtggaga-ctgctt-tgc----g-gatct 1499

Db  1765 GGTCTGTCCCCACAAGTGCCGCTGTGAGGCCAACGTGGTGGAGTGCTCCAGCCTGAAGCT 1824

Qy  1500 ggcttgccctgaaaagtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagct 1559

Db  1825 CACCAAGATCCCTGAGCGCATCCCCCAGTCCACGGCAGAACTGCGATTGAATAACAATGA 1884

Qy  1560 caacaaaatcccggagcacattccccagtacactgcagagttgcgtctcaataataatga 1619

Db  1885 GATTTCCATCCTGGAGGCCACTGGGATGTTTAAAAAACTTACACATCTGAAGAAAATCAA 1944

Qy  1620 atttaccgtgttggaagccacaggaatctttaagaaacttcctcaattacgtaaaataaa 1679

Db  1945 TCTGAGCAACAACAAGGTGTCAGAAATTGAAGATGGGGCCTTCGAGGGCGCAGCCTCTGT 2004

Qy  1680 ctttagcaacaataagatcacagatattgaggagggagcatttgaaggagcatctggtgt 1739

Db  2005 GAGCGAGCTGCACCTAACTGCCAACCAGCTGGAGTCCATCCGGAGCGGCATGTTCCGGGG 2064

Qy  1740 aaatgaaatacttcttacgagtaatcgtttggaaaatgtgcagcataagatgttcaaggg 1799

Db  2065 TCTGGATGGCTTGAGGACCCTAATGCTGCGGAACAACCGCATCAGCTGCATCCACAACGA 2124

Qy  1800 attggaaagcctcaaaactttgatgttgagaagcaatcgaataacctgtgtggggaatga 1859

Db  2125 CAGCTTCACGGGCCTGCGCAACGTCCGGCTCCTCTCGCTCTACGACAACCAGATCACCAC 2184

Qy  1860 cagtttcataggactcagttctgtgcgtttgctttctttgtatgataatcaaattactac 1919

Db  2185 CGTATCCCCAGGAGCCTTCGACACCCTCCAGTCCCTCTCCACACTGAATCTCCTGGCCAA 2244

Qy  1920 agttgcaccaggggcatttgatactctccattctttatctactctaaacctcttggccaa 1979

Db  2245 CCCTTTCAACTGCAACTGCCAGCTGGCCTGGCTAGGAGGCTGGCTACGGAAGCGCAAGAT 2304

Qy  1980 tccttttaactgtaactgctacctggcttggttgggagagtggctgagaaagaagagaat 2039

Db  2305 CGTGACGGGGAACCCGCGATGCCAGAACCCTGACTTTTTGCGGCAGATTCCCCTGCAGGA 2364

Qy  2040 tgtcacgggaaatcctagatgtcaaaaaccatacttcctgaaagaaatacccatccagga 2099

Db  2365 CGTGGCCTTCCCTGACTTCAGGTGTGAGGAAGGCCAGGAGGAGGGGGGCTGCCTGCCCCG 2424

Qy  2100 tgtggccattcaggacttcacttgtgatgacggaaatgatgacaatagttgctccccact 2159

Db  2425 CCCACAGTGCCCACAGGAGTGCGCCTGCCTGGACACCGTGGTCCGATGCAGCAACAAGCA 2484

Qy  2160 ttctcgctgtcctactgaatgtacttgcttggatacagtcgtccgatgtagcaacaaggg 2219

Db  2485 CCTGCGGGCCCTGCCCAAGGGCATTCCCAAGAATGTCACAGAACTCTATTTGGACGGGAA 2544

Qy  2220 tttgaaggtcttgccgaaaggtattccaagagatgtcacagagttgtatctggatggaaa 2279

Db  2545 CCAGTTCACGCTGGTTCCGGGACAGCTGTCTACCTTCAAGTACCTGCAGCTCGTGGACCT 2604

Qy  2280 ccaatttacactggttcccaaggaactctccaactacaaacatttaacacttatagactt 2339

Db  2605 GAGCAACAACAAGATCAGTTCCTTAAGCAATTCCTCCTTCACCAACATGAGCCAGCTGAC 2664

Qy  2340 aagtaacaacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcct 2399

Db  2665 CACTCTGATCCTCAGCTACAATGCCCTGCAGTGCATCCCGCCTTTGGCCTTCCAGGGAGT 2724

Qy  2400 caccttaattcttagttacaaccgtctgagatgtattcctcctcgcacctttgatggatt 2459

Db  2725 CCGCTCCCTGCGCCTGCTGTCTCTCCACGGCAATGACATCTCCACCCTCCAAGAGGGCAT 2784

Qy  2460 aaagtctcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgc 2519

Db  2785 CTTTGCAGACGTGACCTCCCTGTCTCACCTGGCCATTGGTGCCAACCCCCTATACTGTGA 2844

Qy  2520 tttcaatgatctttctgcattatcacatctagcaattggagccaaccctctttactgtga 2579

Db  2845 CTGCCACCTCCGCTGGCTGTCCAGCTGGGTGAAGACTGGCTACAAGGAACCGGGCATTGC 2904

Qy  2580 ttgtaacatgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgc 2639

Db  2905 TCGTTGTGCTGGGCCCCAGGACATGGAGGGCAAGCTGCTCCTCACCACGCCTGCCAAGAA 2964

Qy  2640 tcgttgtgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaa 2699

Db  2965 GTTTGAATGCCAAGGTCCTCCAACGCTGGCTGTCCAGGCCAAGTGTGATCTCTGCTTGTC 3024

Qy  2700 atttacctgtcaaggtcctgtggatgtcaatattctagctaagtgtaacccctgcctatc 2759

Db  3025 CAGTCCGTGCCAGAACCAGGGCACCTGCCACAACGACCCCCTTGAGGTGTACAGGTGCGC 3084

Qy  2760 aaatccgtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcac 2819

Db  3085 CTGCCCCAGCGGCTATAAGGGTCGAGACTGTGAGGTGTCCCTGAACAGCTGTTCCAGTGG 3144

Qy  2820 ctgtccatatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaa 2879

Db  3145 CCCCTGTGAAAATGGGGGCACCTGCCATGCACAGGAGGGCGAGGATGCCCCGTTCACGTG 3204

Qy  2880 cccatgtaaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtg 2939

Db  3205 CTCCTGTCCCACCGGCTTTGAAGGACCAACCTGTGGGGTGAACACAGATGACTGTGTGGA 3264

Qy  2940 tatttgtgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaaga 2999

Db  3265 TCATGCCTGTGCCAATGGGGGCGTCTGTGTGGATGGTGTGGGCAACTACACCTGCCAGTG 3324

Qy  3000 taatgactgtgaaaataattctacatgtgtcgatggcattaataactacacatgcctttg 3059

Db  3325 CCCCCTGCAGTATGAGGGAAAGGCCTGTGAGCAGCTGGTGGACTTGTGCTCTCCGGATCT 3384

Qy  3060 cccacctgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacct 3119

Db  3385 GAACCCATGTCAACACGAGGCCCAGTGTGTGGGCACCCCGGATGGGCCCAGGTGTGAGTG 3444

Qy  3120 gaacccctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactg 3179

Db  3445 CATGCCAGGTTATGCAGGTGACAACTGCAGTGAGAACCAGGATGACTGCAGGGACCACCG 3504

Qy  3180 cacaccagggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaa 3239

Db  3505 CTGCCAGAATGGGGCCCAGTGTATGGATGAAGTCAACAGCTACTCCTGCCTCTGTGCTGA 3564

Qy  3240 gtgtaaaaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccga 3299

Db  3565 GGGCTACAGTGGACAGCTCTGTGAGATCCCTCC--CCATCTGCCTGCCCCCAAG-AGCCC 3621

Qy  3300 aggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagccc 3359

Db  3622 CTGTGAGGGGACTGAGTGCCAGAATGGGGCCAACTGTGTGGACCAGGGCAACAGGCCTGT 3681

Qy  3360 ctgtgataatttttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaat 3419

Db  3682 GTGCCAGTGCCTCCCAGGCTTCGGTGGCCCTGAGTGTGAGAAGTTGCTCAGTGTCAACTT 3741

Qy  3420 atgtcagtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaattt 3479

Db  3742 TGTGGATCGGGACACTTACCTGCAGTTCACTGACCTGCAAAACTGGCCACGGGCCAACAT 3801

Qy  3480 tataaacaaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacat 3539

Db  3802 CACGTTGCAGGTCTCCACGGCAGAGGACAATGGGATCCTTCTGTACAACGGGGACAACGA 3861

Qy  3540 aacacttcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaaga 3599
```

```
Db   3862 CCACATTGCAGTTGAGCTGTACCAGGGCCATGTGCGTGTCAGCTACGACCCAGGCAGCTA 3921
Qy   3600 ccatatcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctca 3659

Db   3922 CCCCAGCTCTGCCATCTACAGTGCTGAGACGATCAACGATGGGCAATTCCACACCGTTGA 3981
Qy   3660 tccagcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtgga 3719

Db   3982 GCTGGTTGCCTTTGACCAGATGGTGAATCTCTCCATTGATGGCGGGAGCCCCATGACCAT 4041
Qy   3720 actacttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcat 3779

Db   4042 GGACAACTTTGGCAAACATTACACGCTCAACAGCGAGGCGCCACTCTATGTGGGAGGGAT 4101
Qy   3780 cactaacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcat 3839

Db   4102 GCCCGTGGATGTCAACTCAGCTGCCTTCCGCCTGTGGCAGATCCTCAACGGCACCGGCTT 4161
Qy   3840 gccagggaagagtaacgtggcatctctgcgccaggcccctgggcagaacggaaccagctt 3899

Db   4162 CCACGGTTGCATCCGAAACCTGTACATCAACAACGAGCTGCAGGACTTCACCAAGACGCA 4221
Qy   3900 ccacggctgcatccggaacctttacatcaacagtgagctgcaggacttccagaaggtgcc 3959

Db   4222 GATGAAGCCAGGCGTGGTGCCAGGCTGCGAACCCTGCCGCAAGCTCTACTGCCTGCATGG 4281
Qy   3960 gatgcaaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatgg 4019

Db   4282 CATCTGCCAGCCCAATGCCACCCCAGGGCCCATGTGCCACTGCGAGGCTGGCTGGGTGGG 4341
Qy   4020 cacatgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggg 4079

Db   4342 CCTGCACTGTGACCAGCCCGCTGACGGCCCCTGCCATGGCCACAAGTGTGTCCATGGGCA 4401
Qy   4080 gcccctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcac 4139

Db   4402 ATGCGTGCCCCTCGACGCTCTTTCCTACAGCTGCCAGTGCCAGGATGGGTACTCGGGGGC 4461
Qy   4140 ctgcttgcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgt 4199

Db   4462 ACTGTGCAACCAGGCCGGGGCCCTGGCAGAGCCCTGCAGAGGCCTGCAGTGCCTGCATGG 4521
Qy   4200 cctctgtgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatgg 4259

Db   4522 CCACTGCCAGGCCTCAGGCACCAAGGGGGCACACTGTGTGTGTGACCCCGGCTTTTCGGG 4581
Qy   4260 gaagtgcaggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacggg 4319

Db   4582 CGAGCTGTGTGAGCAAGAGTCCGAGTGCCGGGGGGACCCTGTCCGGGACTTTCACCAGGT 4641
Qy   4320 ggacagctgtgatcgagaaatctcttgtcgaggggaaaggataagagattattaccaaaa 4379

Db   4642 CCAGAGGGGCTATGCCATCTGCCAGACCACGCGCCCCCTGTCATGGGTGGAGTGCCGGGG 4701
Qy   4380 gcagcagggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagagg 4439

Db   4702 CTCGTGCCCAGGCCAGGGCTGCTGCCAGGGCCTTCGGCTGAAGCGGAGGAAGTTCACCTT 4761
Qy   4440 tgggtgtgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactcttt 4499

Db   4762 TGAGTGCAGCGATGGGACCTCTTTTGCCGAGGAGGTGGAAAAGCCCACCAAGTGTGGCTG 4821
Qy   4500 cgaatgcactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559

Db   4822 T 4822
Qy   4560 t 4560
```

RESULT 6
LOCUS AB017169 5015 bp mRNA PRI 06-FEB-1999
DEFINITION Homo sapiens mRNA for Slit-3 protein, complete cds.

ACCESSION AB017169
NID g4049588
VERSION AB017169.1 GI:4049588
KEYWORDS slit-3; Slit-3 protein.
SOURCE Homo sapiens fetal tissue_lib:lung cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5015)
AUTHORS Itoh,A. and Sakano,S.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) to the DDBJ/EMBL/GenBank databases. Akira Itoh, Asahi Chemical Industry co.,ltd., Life Science Fundamental Research Laboratory; 2-1, Samejima, Fuji, Shizuoka 416-8501, Japan (E-mail:a8611483@ut.asahi-kasei.co.jp, Tel:+81-545-62-3231, Fax:+81-545-62-3249)
REFERENCE 2 (sites)
AUTHORS Itoh,A., Miyabayashi,T., Ohno,M. and Sakano,S.
TITLE Cloning and expressions of three mammalian homologues of Drosophila slit suggest possible roles for Slit in the formation and maintenance of the nervous system
JOURNAL Brain Res. Mol. Brain Res. 62 (2), 175-186 (1998)
MEDLINE 99033071
FEATURES Location/Qualifiers
source 1. .5015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetal"
/tissue_lib="lung"
gene 264. .4835
/gene="slit-3"
CDS 264. .4835
/gene="slit-3"
/codon_start=1
/product="Slit-3 protein"
/protein_id="BAA35186.1"
/db_xref="PID:d1036172"
/db_xref="PID:g4049589"
/db_xref="GI:4049589"
/translation="MAPGWAGVGAAVRARLALALALASVLSGPPAVACPTKCTCSAAS
VDCHGLGLRAVPRGIPRNAERLDLDRNNITRITKMDFAGLKNLRVLHLEDNQVSVIER
GAFQDLKQLERLRLNKNKLQVLPELLFQSTPKLTRLDLSENQIQGIPRKAFRGITDVK
NLQLDNNHISCIEDGAFRALRDLEILTLNNNNISRILVTSFNHMPKIRTLRLHSNHLY
CDCHLAWLSDWLRQRRTVGQFTLCMAPVHLRGFNVADVQKKEYVCPAPHSEPPSCNAN
SISCPSPCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQYKKLK
RIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLLNANKIN
CLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHLAQNPFVCDCHLKWLAD
YLQDNPIETSGARCSSPRRLANKRISQIKSKKFRCSGSEDYRSRFSSECFMDLVCPEK
CRCEGTIVDCSNQKLVRIPSHLPEYVTDLRLNDNEVSVLEATGIFKKLPNLRKINLSN
NKIKEVREGAFDGAASVQELMLTGNQLETVHGRVFRGLSGLKTLMLRSNLIGCVSNDT
FAGLSSVRLLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLAWLGKWLRKRR
IVSGNPRCQKPFFLKEIPIQDVAIQDFTCDGNEESSCQLSPRCPEQCTCMETVVRCSN
KGLRALPRGMPKDVTELYLEGNHLTAVPRELSALRHLTLIDLSNNSISMLTNYTFSNM
SHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSLSHLALGT
NPLHCDCSLRWLSEWVKAGYKEPGIARCSSPEPMADRLLLTTPTHRFQCKGPVDINIV
AKCNACLSSPCKNNGTCTQDPVELYRCACPYSYKGKDCTVPINTCIQNPCQHGGTCHL
SDSHKDGFSCSCPLGFEGQRCEINPDDCEDNDCENNATCVDGINNYVCICPPNYTGEL
CDEVIDHCVPELNLCQHEAKCIPLDKGFSCECVPGYSGKLCETDNDCVAHKCRHGAQ
CVDTINGYTCTCPQGFSGPFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQQEPTCRCP
PGFAGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYKGDNDPL
ALELYQGHVRLVYDSLSSPPTTVYSVETVNDGQFHSVELVTLNQTLNLVVDKGTPKSL
GKLQKQPAVGINSPLYLGGIPTSTGLSALRQGTDRPLGGFHGCIHEVRINNELQDFKA
LPPQSLGVSPGCKSCTVCKHGLCRSVEKDSVVCECRPGWTGPLCDQEARDPCLGHRCH
HGKCVATGTSYMCKCAEGYGGDLCDNKNDSANACSAFKCHHGQCHISDQGEPYCLCQP
GFSGEHCQQENPCLGQVVREVIRRQKGYASCATASKVPIMECRGGCGPQCCQPTRSKR
RKYVFQCTDGSSFVEEVERHLECGCLACS"
BASE COUNT 1119 a 1611 c 1376 g 909 t
ORIGIN

Query Match 25.3%; Score 1202; DB 29; Length 5015;
Best Local Similarity 64.6%; Pred. No. 0.00e+00;
Matches 2894; Conservative 0; Mismatches 1566; Indels 21; Gaps 13;

```
Db   360 GCCTGCCCCACCAAGTGTACCTGCTCCGCTGCCAGCGTGGACTGCCACGGGCTGGGCCTC 419
Qy    79 gcgtgcccggcgcagtgctcttgctcgggcagcacagtggactgtcacgggctggcgctg 138

Db   420 CGCGCGGTTCCTCGGGGCATCCCCCGCAACGCTGAGCGCCTTGACCTGGACAGAAATAAT 479
Qy   139 cgcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatggaaataac 198

Db   480 ATCACCAGGATCACCAAGATGGACTTCGCTGGGCTCAAGAACCTCCGAGTCTTGCATCTG 539
Qy   199 atcacaagaattacgaagacagattttgctggtcttagacatctaagagttcttcagctt 258

Db   540 GAAGACAACCAGGTCAGCGTCATCGAGAGAGGCGCCTTCCAGGACCTGAAGCAGCTAGAG 599
Qy   259 atggagaataagattagcaccattgaaagaggagcattccaggatcttaaagaactagag 318

Db   600 CGACTGCGCCTGAACAAGAATAAGCTGCAAGTCCTTCCAGAATTGCTTTTCCAGAGCACG 659
Qy   319 agactgcgtttaaacagaaatcaccttcagctgtttcctgagttgctgtttcttgggact 378

Db   660 CCGAAGCTCACCAGACTAGATTTGAGTGAAAACCAGATCCAGGGGATCCCGAGGAAGGCG 719
Qy   379 gcgaagctatacaggcttgatctcagtgaaaaccaaattcaggcaatcccaaggaaagct 438

Db   720 TTCCGCGGCATCACCGATGTGAAGAACCTGCAACTGGACAACAACCACATCAGCTGCATT 779
Qy   439 ttccgtggggcagttgacataaaaaatttgcaactggattacaaccagatcagctgtatt 498

Db   780 GAAGATGGAGCCTTCCGAGCGCTGCGCGATTTGGAGATCCTTACCCTCAACAACAACAAC 839
Qy   499 gaagatggggcattcagggctctccgggacctggaagtgctcactctcaacaataacaac 558

Db   840 ATCAGTCGCATCCTGGTCACCAGCTTCAACCACATGCCGAAGATCCGAACTCTGCGCCTC 899
Qy   559 attactagactttctgtggcaagtttcaaccatatgcctaaacttaggacttttcgactg 618

Db   900 CACTCCAACCACCTGTACTGCGACTGCCACCTGGCCTGGCTCTCGGATTGGCTGCGACAG 959
Qy   619 cattcaaacaacctgtattgtgactgccacctggcctggctctccgactggcttcgcaaa 678

Db   960 CGACGGACAGTTGGCCAGTTCACACTCTGCATGGCTCCTGTGCATTTGAGGGGCTTCAAC 1019
Qy   679 aggcctcgggttggtctgtacactcagtgtatgggcccctcccacctgagaggccataat 738

Db  1020 GTGGCGGATGTGCAGAAGAAGGAGTACGTGTGCCCAGCCCCCCACTCGGAGCCCCCATCC 1079
Qy   739 gtagccgaggttcaaaaacgagaatttgtctgcagtgatgaggaagaaggtcaccagtca 798

Db  1080 TGCAATGC--CAACTCC-A----TCT-C-CTGCCCTTCGCCCTGCACGTGCAGCAATAAC 1130
Qy   799 tttatggctccttcttgtagtgttttgcactgccctgccgcctgtacctgtagcaacaat 858

Db  1131 ATCGTGGACTGTCGAGGAAAGGGCTTGATGGAGATTCCTGCCAACTTGCCGGAGGGCATC 1190
Qy   859 atcgtagactgtcgtgggaaaggtctcactgagatccccacaaatcttccagagaccatc 918

Db  1191 GTCGAAATACGCCTAGAACAGAACTCCATCAAAGCCATCCCTGCAGGAGCCTTCACCCAG 1250
Qy   919 acagaaatacgtttggaacagaacacaatcaaagtcatccctcctggagctttctcacca 978

Db  1251 TACAAGAAACTGAAGCGAATAGACATCAGCAAGAATCAGATATCGGATATTGCTCCAGAT 1310
Qy   979 tataaaaagcttagacgaattgacctgagcaataatcagatctctgaacttgcaccagat 1038

Db  1311 GCCTTCCAGGGCCTGAAATCACTCACATCGCTGGTCCTGTATGGGAACAAGATCACCGAG 1370
Qy  1039 gctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcacagaa 1098

Db  1371 ATTGCCAAGGGACTGTTTGATGGGCTGGTGTCCCTACAGCTGCTCCTCCTCAATGCCAAC 1430
Qy  1099 ctccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgccaac 1158

Db  1431 AAGATCAACTGCCTGCGGGTGAACACGTTTCAGGACCTGCAGAACCTCAACTTGCTCTCC 1490

Qy  1159 aagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctctcc 1218

Db  1491 CTGTATGACAACAAGCTGCAGACCATCAGCAAGGGGCTCTTCGCCCCTCTGCAGTCCATC 1550
Qy  1219 ctatatgacaacaagcttcagaccatcgccaaggggacctttttcacctcttcgggccatt 1278

Db  1551 CAGACACTCCACTTAGCCCAAAACCCATTTGTGTGCGACTGCCACTTGAAGTGGCTGGCC 1610
Qy  1279 caaactatgcatttggcccagaaccccttttatttgtgactgccatctcaagtggctagcg 1338

Db  1611 GACTACCTCCAGGACAACCCCATCGAGACAAGCGGGGCCCGCTGCAGCAGCCCGCGCCGA 1670
Qy  1339 gattatctccataccaacccgattgagaccagtggtgcccgttgcaccagcccccgccgc 1398

Db  1671 CTCGCCAACAAGCGCATCAGCCAGATCAAGAGCAAGAAGTTCCGCTGCTCAGGCTCCGAG 1730
Qy  1399 ctggcaaacaaaagaattggacagatcaaaagcaagaaattccgttgttcaggtacagaa 1458

Db  1731 GATTACCGCAGCAGGTTCAGCAGCGAGTGCTTCATGGACCTCGTGTGCCCCGAGAAGTGT 1790
Qy  1459 gattatcgatcaaaattaagtggagactgctttgcggatctggcttgccctgaaaagtgt 1518

Db  1791 CGCTGTGAGGGCACGATTGTGGACTGCTCCAACCAGAAGCTGGTCCGCATCCCAAGCCAC 1850
Qy  1519 cgctgtgaaggaaccacagtagattgctctaatcaaaagctcaacaaaatcccggagcac 1578

Db  1851 CTCCCTGAATATGTCACCGACCTGCGACTGAATGACAATGAGGTATCTGTTCTGGAGGCC 1910
Qy  1579 attccccagtacactgcagagttgcgtctcaataataatgaatttaccgtgttggaagcc 1638

Db  1911 ACTGGCATCTTCAAGAAGTTGCCCAACCTGCGGAAAATAAATCTGAGTAACAATAAGATC 1970
Qy  1639 acaggaatctttaagaaacttcctcaattacgtaaaataaactttagcaacaataagatc 1698

Db  1971 AAGGAGGTGCGAGAGGGAGCTTTCGATGGAGCAGCCAGCGTGCAGGAGCTGATGCTGACA 2030
Qy  1699 acagatattgaggagggagcatttgaaggagcatctggtgtaaatgaaatacttcttacg 1758

Db  2031 GGGAACCAGCTGGAGACCGTGCACGGGCGCGTGTTCCGTGGCCTCAGTGGCCTCAAAACC 2090
Qy  1759 agtaatcgtttggaaaatgtgcagcataagatgttcaagggattggaaagcctcaaaact 1818

Db  2091 TTGATGCTGAGGAGTAACTTGATCGGCTGTGTGAGTAATGACACCTTTGCCGGCCTGAGT 2150
Qy  1819 ttgatgttgagaagcaatcgaataacctgtgtggggaatgacagtttcataggactcagt 1878

Db  2151 TCGGTGAGACTGCTGTCCCTCTATGACAATCGGATCACCACCATCACCCCTGGGGCCTTC 2210
Qy  1879 tctgtgcgtttgctttctttgtatgataatcaaattactacagttgcaccaggggcattt 1938

Db  2211 ACCACGCTTGTCTCCCTGTCCACCATAAACCTCCTGTCCAACCCCTTCAACTGCAACTGC 2270
Qy  1939 gatactctccattctttatctactctaaacctcttggccaatccttttaactgtaactgc 1998

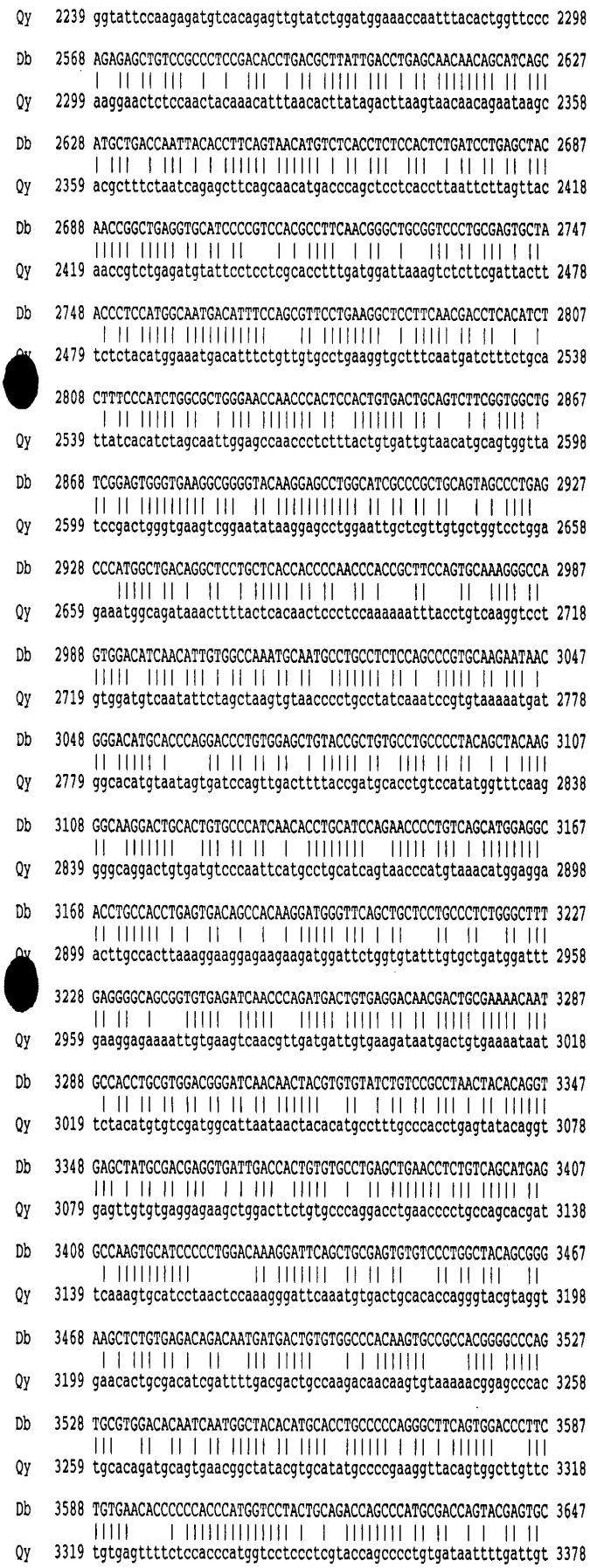
Db  2271 CACCTGGCCTGGCTCGGCAAGTGGTTGAGGAAGAGGCGGATCGTCAGTGGGAACCCTAGG 2330
Qy  1999 tacctggcttggttgggagagtggctgagaaagaagagaattgtcacgggaaatcctaga 2058

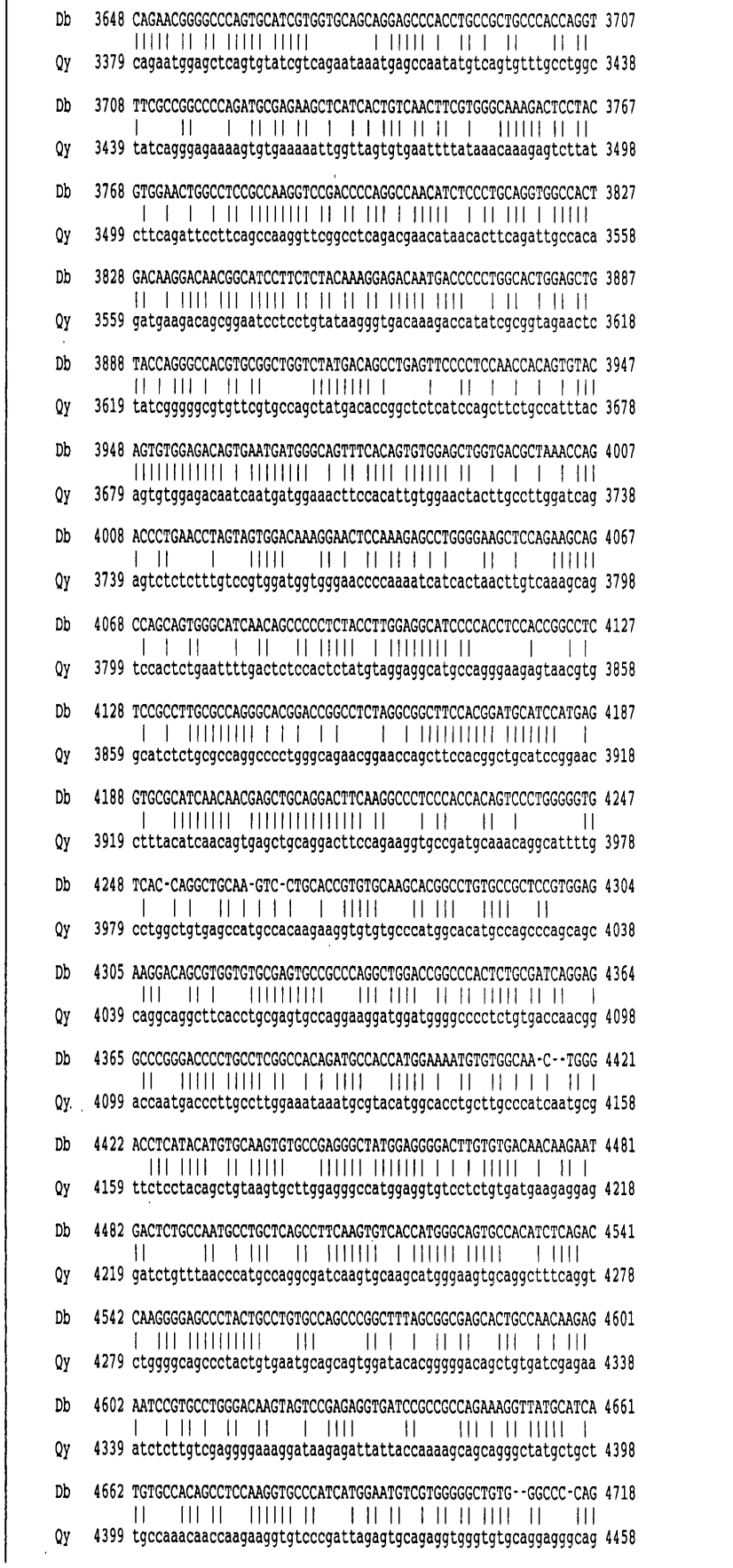
Db  2331 TGCCAGAAGCCATTTTTCCTCAAGGAGATTCCCATCCAGGATGTGGCCATCCAGGACTTC 2390
Qy  2059 tgtcaaaaaccatacttcctgaaagaaatacccatccaggatgtggccattcaggacttc 2118

Db  2391 ACCTGTGATGGC---AACGAGGAGAGTAGCTGCCAGCTGAGCCCGCGCTGCCCGGAGCAG 2447
Qy  2119 acttgtgatgacggaaatgatgacaatagttgctccccactttctcgctgtcctactgaa 2178

Db  2448 TGCACCTGTATGGAGACAGTGGTGCGATGCAGCAACAAGGGGCTCCGCGCCCTCCCCAGA 2507
Qy  2179 tgtacttgcttggatacagtcgtccgatgtagcaacaagggtttgaaggtcttgccgaaa 2238

Db  2508 GGCATGCCCAAGGATGTGACCGAGCTGTACCTGGAAGGAAACCACCTAACAGCCGTGCCC 2567
```


Qy 2239 ggtattccaagagatgtcacagagttgtatctggatggaaaccaatttacactggttccc 2298
Db 2568 AGAGAGCTGTCCGCCCTCCGACACCTGACGCTTATTGACCTGAGCAACAACAGCATCAGC 2627
Qy 2299 aaggaactctccaactacaaacatttaacacttatagacttaagtaacaacagaataagc 2358
Db 2628 ATGCTGACCAATTACACCTTCAGTAACATGTCTCACCTCTCCACTCTGATCCTGAGCTAC 2687
Qy 2359 acgctttctaatcagagcttcagcaacatgacccagctcctcaccttaattcttagttac 2418
Db 2688 AACCGGCTGAGGTGCATCCCCGTCCACGCCTTCAACGGGCTGCGGTCCCTGCGAGTGCTA 2747
Qy 2419 aaccgtctgagatgtattcctcctcgcacctttgatggattaaagtctcttcgattactt 2478
Db 2748 ACCCTCCATGGCAATGACATTTCCAGCGTTCCTGAAGGCTCCTTCAACGACCTCACATCT 2807
Qy 2479 tctctacatggaaatgacatttctgttgtgcctgaaggtgctttcaatgatctttctgca 2538
2808 CTTTCCCATCTGGCGCTGGGAACCAACCCACTCCACTGTGACTGCAGTCTTCGGTGGCTG 2867
Qy 2539 ttatcacatctagcaattggagccaaccctctttactgtgattgtaacatgcagtggtta 2598
Db 2868 TCGGAGTGGGTGAAGGCGGGGTACAAGGAGCCTGGCATCGCCCGCTGCAGTAGCCCTGAG 2927
Qy 2599 tccgactgggtgaagtcggaatataaggagcctggaattgctcgttgtgctggtcctgga 2658
Db 2928 CCCATGGCTGACAGGCTCCTGCTCACCACCCCAACCCACCGCTTCCAGTGCAAAGGGCCA 2987
Qy 2659 gaaatggcagataaacttttactcacaactccctccaaaaaatttacctgtcaaggtcct 2718
Db 2988 GTGGACATCAACATTGTGGCCAAATGCAATGCCTGCCTCTCCAGCCCGTGCAAGAATAAC 3047
Qy 2719 gtggatgtcaatattctagctaagtgtaacccctgcctatcaaatccgtgtaaaaatgat 2778
Db 3048 GGGACATGCACCCAGGACCCTGTGGAGCTGTACCGCTGTGCCTGCCCCTACAGCTACAAG 3107
Qy 2779 ggcacatgtaatagtgatccagttgacttttaccgatgcacctgtccatatggtttcaag 2838
Db 3108 GGCAAGGACTGCACTGTGCCCATCAACACCTGCATCCAGAACCCCTGTCAGCATGGAGGC 3167
Qy 2839 gggcaggactgtgatgtcccaattcatgcctgcatcagtaacccatgtaaacatggagga 2898
Db 3168 ACCTGCCACCTGAGTGACAGCCACAAGGATGGGTTCAGCTGCTCCTGCCCTCTGGGCTTT 3227
Qy 2899 acttgccacttaaaggaaggagaagaagatggattctggtgtatttgtgctgatggattt 2958
3228 GAGGGGCAGCGGTGTGAGATCAACCCAGATGACTGTGAGGACAACGACTGCGAAAACAAT 3287
Qy 2959 gaaggagaaaattgtgaagtcaacgttgatgattgtgaagataatgactgtgaaaataat 3018
Db 3288 GCCACCTGCGTGGACGGGATCAACAACTACGTGTGTATCTGTCCGCCTAACTACACAGGT 3347
Qy 3019 tctacatgtgtcgatggcattaataactacacatgcctttgcccacctgagtatacaggt 3078
Db 3348 GAGCTATGCGACGAGGTGATTGACCACTGTGTGCCTGAGCTGAACCTCTGTCAGCATGAG 3407
Qy 3079 gagttgtgtgaggagaagctggacttctgtgcccaggacctgaacccctgccagcacgat 3138
Db 3408 GCCAAGTGCATCCCCCTGGACAAAGGATTCAGCTGCGAGTGTGTCCCTGGCTACAGCGGG 3467
Qy 3139 tcaaagtgcatcctaactccaaagggattcaaatgtgactgcacaccagggtacgtaggt 3198
Db 3468 AAGCTCTGTGAGACAGACAATGATGACTGTGTGGCCCACAAGTGCCGCCACGGGGCCCAG 3527
Qy 3199 gaacactgcgacatcgattttgacgactgccaagacaacaagtgtaaaaacggagcccac 3258
Db 3528 TGCGTGGACACAATCAATGGCTACACATGCACCTGCCCCCAGGGCTTCAGTGGACCCTTC 3587
Qy 3259 tgcacagatgcagtgaacggctatacgtgcatatgccccgaaggttacagtggcttgttc 3318
Db 3588 TGTGAACACCCCCCACCCATGGTCCTACTGCAGACCAGCCCATGCGACCAGTACGAGTGC 3647
Qy 3319 tgtgagttttctccacccatggtcctccctcgtaccagcccctgtgataatttgattgt 3378


Db 3648 CAGAACGGGGCCCAGTGCATCGTGGTGCAGCAGGAGCCCACCTGCCGCTGCCCACCAGGT 3707
Qy 3379 cagaatggagctcagtgtatcgtcagaataaatgagccaatatgtcagtgtttgcctggc 3438
Db 3708 TTCGCCGGCCCCAGATGCGAGAAGCTCATCACTGTCAACTTCGTGGGCAAAGACTCCTAC 3767
Qy 3439 tatcagggagaaaagtgtgaaaaattggttagtgtgaattttataaacaaagagtcttat 3498
Db 3768 GTGGAACTGGCCTCCGCCAAGGTCCGACCCCAGGCCAACATCTCCCTGCAGGTGGCCACT 3827
Qy 3499 cttcagattccttcagccaaggttcggcctcagacgaacataacacttcagattgccaca 3558
Db 3828 GACAAGGACAACGGCATCCTTCTCTACAAAGGAGACAATGACCCCCTGGCACTGGAGCTG 3887
Qy 3559 gatgaagacagcggaatcctcctgtataagggtgacaaagaccatatcgcggtagaactc 3618
Db 3888 TACCAGGGCCACGTGCGGCTGGTCTATGACAGCCTGAGTTCCCCTCCAACCACAGTGTAC 3947
Qy 3619 tatcgggggcgtgttcgtgccagctatgacaccggctctcatccagcttctgccatttac 3678
Db 3948 AGTGTGGAGACAGTGAATGATGGGCAGTTTCACAGTGTGGAGCTGGTGACGCTAAACCAG 4007
Qy 3679 agtgtggagacaatcaatgatggaaacttccacattgtggaactacttgccttggatcag 3738
Db 4008 ACCCTGAACCTAGTAGTGGACAAAGGAACTCCAAAGAGCCTGGGGAAGCTCCAGAAGCAG 4067
Qy 3739 agtctctctttgtccgtggatggtgggaaccccaaaatcatcactaacttgtcaaagcag 3798
Db 4068 CCAGCAGTGGGCATCAACAGCCCCCTCTACCTTGGAGGCATCCCCACCTCCACCGGCCTC 4127
Qy 3799 tccactctgaattttgactctccactctatgtaggaggcatgccagggaagagtaacgtg 3858
Db 4128 TCCGCCTTGCGCCAGGGCACGGACCGGCCTCTAGGCGGCTTCCACGGATGCATCCATGAG 4187
Qy 3859 gcatctctgcgccaggcccctgggcagaacggaaccagcttccacggctgcatccggaac 3918
Db 4188 GTGCGCATCAACAACGAGCTGCAGGACTTCAAGGCCCTCCCACCACAGTCCCTGGGGGTG 4247
Qy 3919 ctttacatcaacagtgagctgcaggacttccagaaggtgccgatgcaaacaggcattttg 3978
Db 4248 TCAC-CAGGCTGCAA-GTC-CTGCACCGTGTGCAAGCACGGCCTGTGCCGCTCCGTGGAG 4304
Qy 3979 cctggctgtgagccatgccacaagaaggtgtgtgcccatggcacatgccagcccagcagc 4038
Db 4305 AAGGACAGCGTGGTGTGCGAGTGCCGCCCAGGCTGGACCGGCCCACTCTGCGATCAGGAG 4364
Qy 4039 caggcaggcttcacctgcgagtgccaggaaggatggatggggcccctctgtgaccaacgg 4098
Db 4365 GCCCGGGACCCCTGCCTCGGCCACAGATGCCACCATGGAAAATGTGTGGCAA-C--TGGG 4421
Qy 4099 accaatgacccttgccttggaaataaatgcgtacatggcacctgcttgcccatcaatgcg 4158
Db 4422 ACCTCATACATGTGCAAGTGTGCCGAGGGCTATGGAGGGGACTTGTGTGACAACAAGAAT 4481
Qy 4159 ttctcctacagctgtaagtgcttggagggccatggaggtgtcctctgtgatgaagaggag 4218
Db 4482 GACTCTGCCAATGCCTGCTCAGCCTTCAAGTGTCACCATGGGCAGTGCCACATCTCAGAC 4541
Qy 4219 gatctgtttaacccatgccaggcgatcaagtgcaagcatgggaagtgcaggctttcaggt 4278
Db 4542 CAAGGGGAGCCCTACTGCCTGTGCCAGCCCGGCTTTAGCGGCGAGCACTGCCAACAAGAG 4601
Qy 4279 ctggggcagccctactgtgaatgcagcagtggatacacgggggacagctgtgatcgagaa 4338
Db 4602 AATCCGTGCCTGGGACAAGTAGTCCGAGAGGTGATCCGCCGCCACAGAAAGGTTATGCATCA 4661
Qy 4339 atctcttgtcgaggggaaaggataagagattattaccaaaagcagcagggctatgctgct 4398
Db 4662 TGTGCCACAGCCTCCAAGGTGCCCATCATGGAATGTCGTGGGGGCTGTG--GGCCC-CAG 4718
Qy 4399 tgccaaacaaccaagaaggtgtcccgattagagtgcagaggtgggtgtgcaggagggcag 4458

```
Db   4719 TGCTGCCAGCCCACCCGCAGCAAGCGGCGGAAATACGTCTTCCAGTGCACGGACGGCTCC 4778
          |||||    ||     | ||||||||||||||||||   ||| | ||||| |||||||||
Qy   4459 tgctgtggaccgctgaggagcaagcggcggaaatactctttcgaatgcactgacggctcc 4518

Db   4779 TCGTTTGTAGAAGAGGTGGAGAGACACTTAGAGTGCGGCTG 4819
          || ||||| || ||||| |||| |    |  ||||||||||
Qy   4519 tcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559


RESULT    7
LOCUS       AB011530     4950 bp    mRNA            ROD       22-AUG-1998
DEFINITION  Rattus norvegicus mRNA for MEGF4, complete cds.
ACCESSION   AB011530
NID         g3449289
VERSION     AB011530.1  GI:3449289
KEYWORDS    MEGF4.
SOURCE      Rattus norvegicus (strain:Sprague-Dawley) adult brain cDNA to mRNA,
            clone_lib:pSPORT 1.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
  REFERENCE 1  (bases 1 to 4950)
  AUTHORS   Nakayama,M., Nakajima,D. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1998) to the DDBJ/EMBL/GenBank databases. Manabu
            Nakayama, Kazusa DNA Research Institute, Laboratory of DNA
            technology; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:nmanabu@kazusa.or.jp, Tel:+81-438-52-3915,
            Fax:+81-438-52-3914)
REFERENCE   2  (sites)
  AUTHORS   Nakayama,M., Nakajima,D., Nagase,T., Nomura,N., Seki,N. and
            Ohara,O.
  TITLE     Identification of high-molecular-weight proteins with multiple
            EGF-like motifs by motif-trap screening
  JOURNAL   Genomics 51 (1), 27-34 (1998)
  MEDLINE   98360089
FEATURES             Location/Qualifiers
     source          1. .4950
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone_lib="pSPORT 1"
                     /dev_stage="adult"
                     /tissue_type="brain"
     gene            160. .4755
                     /gene="MEGF4"
     CDS             160. .4755
                     /gene="MEGF4"
                     /note="rat homologue of Drosophila slit protein; rat
                     slit1"
                     /codon_start=1
                     /product="MEGF4"
                     /protein_id="BAA32460.1"
                     /db_xref="PID:d1033423"
                     /db_xref="PID:g3449290"
                     /db_xref="GI:3449290"
                     /translation="MALTPQRGSSSGLSRPELWLLLWAAAWRLGATACPALCTCTGTT
                     VDCHGTGLQAIPKNIPRNTERLELNGNNITRIHKNDFAGLKQLRVLQLMENQIGAVER
                     GAFDDMKELERLRLNRNQLQVLPELLFQNNQALSRLDLSENSLQAVPRKAFRGATDLK
                     NLQLDKNQISCIEEGAFRALRGLEVLTLNNNNITTIPVSSFNHMPKLRTFRLHSNHLF
                     CDCHLAWLSQWLRQRPTIGLFTQCSGPASLRGLNVAEVQKSEFSCSGQGEAAQVPACT
                     LSSGSCPAMCSCSNGIVDCRGKGLTAIPANLPETMTEIRLELNGIKSIPPGAFSPYRK
                     LRRIDLSNNQIAEIAPDAFQGLRSLNSLVLYGNKITDLPRGVFGGLYTLQLLLLNANK
                     INCIRPDAFQDLQNLSLLSLYDNKIQSLAKGTFTSLRAIQTLHLAQNPFICDCNLKWL
                     ADFLRTNPIETTGARCASPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYHLNSECT
                     SDVACPHKCRCEASVVECSGLKLSKIPERIPQSTTELRLNNNEISILEATGLFKKLSH
                     LKKINLSNNKVSEIEDGTFEGATSVSELHLTANQLESVRSGMFRGLDGLRTLMLRNNR
                     ISCIHNDSFTGLRNVRLLSLYDNHITTISPGAFDTLQALSTLNLLANPFNCNCQLAWL
                     GDWLRKRKIVTGNPRCQNPDFLRQIPLQDVAFPDFRCEEGQEEVGCLPRPQCPQECAC
                     LDTVVRCSNKHLQALPKGIPKNVTELYLDGNQFTLVPGQLSTFKYLQLVDLSNNKISS
                     LSNSSFTNMSQLTTLILSYNALQCIPPLAFQGLRSLRLLSLHGNDVSTLQEGIFADVT
                     SLSHLAIGANPLYCDCHLRWLSSWVKTGYKEPGIARCAGPPEMEGKLLLTTPAKKFEC
                     QGPPSLAVQAKCDPCLSSPCQNQGTCHNDPLEVYRCTCPSGYKGRNCEVSLDSCSSNP
                     CGNGGTCHAQEGEDAGFTCSCPSGFEGLTCGMNTDDCVKHDCVNGGVCVDGIGNYTCQ
                     CPLQYTGRACEQLVDFCSPDLNPCQHEAQCVGTPEGPRCECVPGYTGDNCSKNQDDCK
                     DHQCQNGAQCVDEINSYACLCAEGYSGQLCEIPPAPRNSCEGTECQNGANCVDQGSRP
                     VCQCLPGFGGPECEKLLSVNFVDRDTYLQFTDLQNWPRANITLQVSTAEDNGILLYNG
                     DNDHIAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVTFDQMVNLSIDGG
                     SPMTMDNFGKHYTLNSEAPLYVGGMPVDVNSAAFRLWQILNGTSFHGCIRNLYINNEL
                     QDFTKTQMKPGVVPGCEPCRKLYCLHGICQPNATPGPVCHCEAGWGGLHCDQPVDGPC
                     HGHKCVHGKCVPLDALAYSCQCQDGYSGALCNQVGAVAEPCGGLQCLHGHCQASATRG
                     AHCVCSPGFSGELCEQESECRGDPVRDFHRVQRGYAICQTTRPLSWVECRGACPGQGC
                     CQGLRLKRRKLTFECSDGTSFAEEVEKPTKCGCAPCA"
     polyA_site      4950
                     /note="15 a nucleotides"
BASE COUNT     1076 a   1501 c   1398 g    975 t
ORIGIN

  Query Match          24.7%;  Score 1177;  DB 32;  Length 4950;
  Best Local Similarity  64.9%;  Pred. No. 0.00e+00;
  Matches  2928;  Conservative    0;  Mismatches 1553;  Indels  33;  Gaps  23;

Db    255 GGCGTGCCCCGCCCTCTGCACCTGCACCGGCACTACAGTGGACTGCCACGGCACGGGTTT 314
          |||||||| || |  ||| | ||| | ||||  |||||||||||| |||||   ||   |
Qy     78 ggcgtgcccggcgcagtgctcttgctcgggcagcacagtggactgtcacgggctggcgct 137

Db    315 GCAGGCCATCCCCAAGAACATCCCACGGAACACCGAGCGCCTGGAACTCAATGGTAACAA 374
          ||    | | |||| ||| ||||| || |||||||||  | |||||  | ||||| || ||
Qy    138 gcgcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatggaaataa 197

Db    375 TATCACTCGGATCCATAAGAATGACTTTGCTGGGCTCAAGCAGCTTCGAGTGTTGCAGCT 434
          |||||  | ||    ||||  || |||||||| || |  || ||  ||||  | |||||
Qy    198 catcacaagaattacgaagacagattttgctggtcttagacatctaagagttcttcagct 257

Db    435 GATGGAGAACCAGATTGGAGCTGTGGAACGGGGAGCTTTTGATGACATGAAGGAGTTGGA 494
          ||||||||  |||||  |  | ||| | ||||| ||  | ||  | || ||  | ||
Qy    258 tatggagaataagattagcaccattgaaagaggagcattccaggatcttaaagaactaga 317

Db    495 GCGGCTACGACTGAACCGCAACCAGCTGCAAGTGTTGCCTGAACTACTGTTCCAGAACAA 554
          | | ||  ||  | ||| | || || || ||  |||| |||||  | ||||| |     |
Qy    318 gagactgcgtttaaacagaaatcaccttcagctgtttcctgagttgctgtttcttgggac 377

Db    555 CCAGGCTCTGTCCAGACTGGACCTGAGTGAGAATTCTCTCCAGGCTGTGCCCAGAAAGGC 614
            |   || | ||| || || || ||||| ||     | |||||  | || || || ||
Qy    378 tgcgaagctatacaggcttgatctcagtgaaaaccaaattcaggcaatcccaaggaaagc 437

Db    615 TTTCCGAGGAGCCACAGACCTCAAAAATCTACAGCTGGACAAGAACCAGATCAGCTGCAT 674
          ||||||  || ||    ||| | |||||| | || |||||  | ||||||||||||| ||
Qy    438 tttccgtggggcagttgacataaaaaatttgcaactggattacaaccagatcagctgtat 497

Db    675 CGAGGAAGGGGCCTTCCGTGCCCTACGGGGACTGGAGGTTCTGACCCTGAACAACAACAA 734
          || || ||||| ||| | || || ||||  ||||| || || || || ||||| |||||
Qy    498 tgaagatggggcattcagggctctccgggacctggaagtgctcactctcaacaataacaa 557

Db    735 CATCACCACCATCCCTGTGTCCAGTTTCAACCATATGCCCAAGCTTCGGACCTTCCGGCT 794
          ||| || |   |  ||||| | |||||||||||||||| || ||| |||| || || ||
Qy    558 cattactagactttctgtggcaagtttcaaccatatgcctaaacttaggacttttcgact 617

Db    795 GCACTCCAACCACCTGTTCTGTGACTGTCACCTGGCCTGGCTCTCGCAGTGGCTGAGACA 854
          ||| || ||| ||||||  |||||||| ||||||||||||||||||  | |||||  | |
Qy    618 gcattcaaacaacctgtattgtgactgccacctggcctggctctccgactggcttcgcaa 677

Db    855 GAGGCCCACCATCGGGCTCTTCACCCAGTGCTCTGGGCCCGCCAGCCTTCGGGGCCTCAA 914
          |||||     | || || | ||| |||||    || ||| ||  |||  | ||||  ||
Qy    678 aaggcctcgggttggtctgtacactcagtgtatgggcccctcccacctgagaggccataa 737

Db    915 TGTGGCAGAGGTGCAAAAGAGTGAGTTCAGCTGCTCAGGCCAGGGGGA-GG-CTGCACAA 972
          ||| || ||||| |||||  | || ||   ||||   |  |||  || || |  ||
Qy    738 tgtagccgaggttcaaaaacgagaatttgtctgcagtgatgaggaagaaggtcaccagtc 797

Db    973 GTACCTGCCTGCACCCTCTCATCCGGTTC-CTGCCCAGCCATGTGCAGCTGTAGCAATGG 1031
          |   || || |  | | |  |    | | |||||| |||   || | |||||||||
Qy    798 atttatggctccttcttgtagtgttttgcactgccctgccgcctgtacctgtagcaacaa 857

Db   1032 CATCGTGGACTGCCGTGGCAAAGGCCTCACTGCCATCCCTGCTAACCTTCCGGAGACGAT 1091
          ||||| ||||| ||||| ||||| |||||||  |||||  | || ||||| ||||| ||
```

```
Qy   858 tatcgtagactgtcgtgggaaaggtctcactgagatccccacaaatcttccagagaccat 917

Db  1092 GACCGAGATCCGCCTGGAGCTGAACGGTATCAAATCCATCCCACCAGGAGCCTTCTCTCC 1151

Qy   918 cacagaaatacgtttggaacagaacacaatcaaagtcatccctcctggagctttctcacc 977

Db  1152 CTACAGAAAGCTGCGCAGGATAGACCTGAGTAACAACCAGATTGCTGAGATTGCACCCGA 1211

Qy   978 atataaaaagcttagacgaattgacctgagcaataatcagatctctgaacttgcaccaga 1037

Db  1212 TGCCTTCCAAGGCCTCCGCTCCTTGAATTCCCTGGTCCTCTATGGCAACAAGATCACAGA 1271

Qy  1038 tgctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcacaga 1097

Db  1272 CTTGCCCCGAGGTGTATTTGGAGGCCTGTACACCCTCCAGCTGCTGCTTCTGAATGCCAA 1331

Qy  1098 actccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgccaa 1157

Db  1332 CAAGATCAACTGCATCCGGCCTGATGCCTTCCAGGACCTGCAGAACCTCTCTCTGCTCTC 1391

Qy  1158 caagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctctc 1217

Db  1392 CCTATATGACAACAAGATCCAGAGTCTCGCCAAGGGCACCTTCACCTCCCTGCGGGCCAT 1451

Qy  1218 cctatatgacaacaagcttcagaccatcgccaaggggaccttttcacctcttcgggccat 1277

Db  1452 CCAGACACTGCATCTGGCCCAGAACCCTTTCATCTGTGACTGTAACCTCAAGTGGCTGGC 1511

Qy  1278 tcaaactatgcatttggcccagaacccctttatttgtgactgccatctcaagtggctagc 1337

Db  1512 AGACTTCTTACGTACCAACCCCATTGAGACAACTGGGGCCCGCTGTGCCAGCCCCCGGCG 1571

Qy  1338 ggattatctccataccaacccgattgagaccagtggtgcccgttgcaccagcccccgccg 1397

Db  1572 CCTTGCCAACAAGCGTATCGGGCAGATCAAAAGCAAGAAGTTCCGGTGCTCAGCCAAAGA 1631

Qy  1398 cctggcaaacaaaagaattggacagatcaaaagcaagaaattccgttgttcaggtacaga 1457

Db  1632 GCAGTACTTCATTCCAGGGACGGAGGATTACCACCTGAACAGCGAGTGTACGAGTGACGT 1691

Qy  1458 --ag-a-tt-at--c---gatcaaa-atta--ag-tggagactgctt-tgcggat--c-t 1499

Db  1692 GGCCTGTCCCCACAAGTGCCGCTGTGAGGCCAGTGTGGTGGAGTGCTCTGGCCTGAAGCT 1751

Qy  1500 ggcttgccctgaaaagtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagct 1559

Db  1752 CTCTAAGATCCCTGAGCGAATCCCCCAGTCCACGACTGAGCTACGACTGAACAACAATGA 1811

Qy  1560 caacaaaatcccggagcacattccccagtacactgcagagttgcgtctcaataataatga 1619

Db  1812 GATTTCAATCCTAGAGGCCACTGGGTTGTTTAAGAAGCTCTCACATCTGAAGAAAATAAA 1871

Qy  1620 atttaccgtgttggaagccacaggaatctttaagaaacttcctcaattacgtaaaataaa 1679

Db  1872 CCTGAGCAACAACAAGGTGTCAGAGATTGAAGACGGGACGTTTGAAGGCGCCACCTCCGT 1931

Qy  1680 ctttagcaacaataagatcacagatattgaggagggagcatttgaaggagcatctggtgt 1739

Db  1932 CAGCGAGCTGCACCTGACCGCCAACCAGCTGGAGTCTGTCAGAAGCGGCATGTTCCGGGG 1991

Qy  1740 aaatgaaatacttcttacgagtaatcgtttggaaaatgtgcagcataagatgttcaaggg 1799

Db  1992 TCTAGATGGCTTGCGGACCCTGATGCTGAGGAACAACCGTATCAGCTGCATCCACAATGA 2051

Qy  1800 attggaaagcctcaaaactttgatgttgagaagcaatcgaataacctgtgtggggaatga 1859

Db  2052 CAGCTTCACTGGGCTGCGTAACGTGCGGCTCCTGTCTCTTTACGACAACCACATCACCAC 2111

Qy  1860 cagtttcataggactcagttctgtgcgtttgctttctttgtatgataatcaaattactac 1919

Db  2112 CATCTCCCCTGGAGCCTTCGACACCCTCCAGGCCCTGTCCACACTAAACCTCCTGGCCAA 2171

Qy  1920 agttgcaccaggggcatttgatactctccattctttatctactctaaacctcttggccaa 1979

Db  2172 CCCATTCAACTGCAACTGCCAGTTGGCCTGGCTGGGGGACTGGCTTCGGAAGAGGAAAAT 2231

Qy  1980 tccttttaactgtaactgctacctggcttggttgggagagtggctgagaaagaagagaat 2039

Db  2232 TGTGACAGGGAACCCTCGCTGTCAGAACCCAGACTTCCTGCGGCAGATTCCTCTGCAAGA 2291

Qy  2040 tgtcacgggaaatcctagatgtcaaaaaccatacttcctgaaagaaatacccatccagga 2099

Db  2292 TGTGGCCTTCCCTGACTTCAGGTGTGAGGAAGGCCAGGAGGAGGTGGGCTGCCTGCCCCG 2351

Qy  2100 tgtggccattcaggacttcacttgtgatgacggaaatgatgacaatagttgctccccact 2159

Db  2352 CCCACAGTGCCCCCAAGAGTGCGCCTGCCTGGACACCGTGGTCCGGTGTAGCAACAAGCA 2411

Qy  2160 ttctcgctgtcctactgaatgtacttgcttggatacagtcgtccgatgtagcaacaaggg 2219

Db  2412 CCTCCAGGCTCTGCCCAAGGGCATCCCCAAGAATGTCACAGAACTCTACTTGGATGGGAA 2471

Qy  2220 tttgaaggtcttgccgaaaggtattccaagagatgtcacagagttgtatctggatggaaa 2279

Db  2472 CCAGTTCACACTGGTTCCGGGACAGCTGTCTACCTTCAAGTATCTGCAGCTTGTCGACCT 2531

Qy  2280 ccaatttacactggttcccaaggaactctccaactacaaacatttaacacttatagactt 2339

Db  2532 GAGCAACAACAAGATCAGCTCCTTAAGCAACTCTTCCTTCACCAACATGAGTCAGCTGAC 2591

Qy  2340 aagtaacaacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcct 2399

Db  2592 CACACTGATCCTCAGCTACAACGCCCTCCAGTGCATCCCCCCTCTGGCCTTCCAAGGACT 2651

Qy  2400 caccttaattcttagttacaaccgtctgagatgtattcctcctcgcaccctttgatggatt 2459

Db  2652 TCGTTCACTGCGCTTGCTGTCTTTGCATGGCAATGATGTCTCCACCCTCCAAGAGGGAAT 2711

Qy  2460 aaagtctcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgc 2519

Db  2712 CTTTGCAGACGTGACGTCTCTGTCTCACCTGGCCATTGGTGCCAACCCCCTGTACTGCGA 2771

Qy  2520 tttcaatgatctttctgcattatcacatctagcaattggagccaaccctctttactgtga 2579

Db  2772 CTGCCACCTCCGATGGCTCTCCAGCTGGGTGAAGACTGGCTACAAGGAACCTGGCATTGC 2831

Qy  2580 ttgtaacatgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgc 2639

Db  2832 CCGCTGTGCTGGACCCCCAGAGATGGAGGGCAAACTGTTGCTCACCACACCTGCCAAGAA 2891

Qy  2640 tcgttgtgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaa 2699

Db  2892 GTTTGAATGCCAAGGTCCCCCCAGCCTGGCGGTTCAGGCCAAGTGTGATCCCTGCTTGTC 2951

Qy  2700 atttacctgtcaaggtcctgtggatgtcaatattctagctaagtgtaaccccctgcctatc 2759

Db  2952 CAGCCCATGCCAGAATCAGGGCACTTGTCACAATGACCCTCTTGAGGTATACAGGTGTAC 3011

Qy  2760 aaatccgtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcac 2819

Db  3012 CTGCCCCAGTGGCTACAAGGGCCGAAACTGTGAGGTGTCTCTGGACAGCTGTTCCAGCAA 3071

Qy  2820 ctgtccatatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaa 2879

Db  3072 CCCATGTGGAAATGGGGGAACCTGCCATGCCCAGGAGGGCGAGGATGCTGGCTTCACGTG 3131

Qy  2880 cccatgtaaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtg 2939

Db  3132 CTCCTGTCCTTCCGGCTTTGAAGGGCTGACGTGTGGAATGAACACAGATGATTGTGTGAA 3191

Qy  2940 tatttgtgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaaga 2999

Db  3192 GCACGACTGTGTCAATGGGGGTGTCTGTGTGGATGGCATAGGGAATTACACATGTCAGTG 3251

Qy  3000 taatgactgtgaaaatgattctacatgtgtcgatggcattaataactacacatgcctttg 3059
```

```
Db  3252 CCCCCTGCAGTACACAGGAAGGGCCTGTGAACAGCTAGTGGACTTCTGCTCTCCGGATCT 3311
Qy  3060 cccacctgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacct 3119

Db  3312 GAACCCATGTCAGCATGAGGCCCAATGTGTTGGCACCCCAGAGGGGCCCAGGTGTGAGTG 3371
Qy  3120 gaacccctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactg 3179

Db  3372 TGTGCCAGGCTACACGGGTGACAATTGCAGCAAAAACCAGGACGACTGCAAGGACCACCA 3431
Qy  3180 cacaccagggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaa 3239

Db  3432 GTGTCAGAACGGGGCTCAGTGTGTGGATGAGATCAACAGCTACGCCTGTCTCTGTGCTGA 3491
Qy  3240 gtgtaaaaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccga 3299

Db  3492 GGGCTACAGCGGACAGCTCTGTGAGAT--C-CCACC--TG--CC-CCCAGG-A--A-CTC 3539
Qy  3300 aggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagccc 3359

Db  3540 CTGTGAAGGGACTGAGTGCCAGAATGGAGCCAACTGTGTAGACCAGGGCAGCCGGCCCGT 3599
Qy  3360 ctgtgataatttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaat 3419

Db  3600 ATGCCAGTGCCTACCAGGCTTTGGAGGTCCCGAATGTGAGAAGTTGCTCAGTGTCAACTT 3659
Qy  3420 atgtcagtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaattt 3479

Db  3660 TGTGGACCGGGACACTTACCTGCAGTTCACTGACCTGCAGAACTGGCCTCGGGCCAACAT 3719
Qy  3480 tataaacaaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacat 3539

Db  3720 CACTCTTCAGGTCTCCACAGCAGAGGACAATGGGATCCTCCTCTACAATGGGGATAATGA 3779
Qy  3540 aacacttcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaaga 3599

Db  3780 CCACATTGCAGTTGAGCTGTACCAGGGCCATGTCCGTGTTAGCTACGACCCAGGCAGCTA 3839
Qy  3600 ccatatcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctca 3659

Db  3840 CCCCAGCTCTGCTATCTACAGTGCTGAAACAATCAACGATGGGCAGTTCCACACAGTTGA 3899
Qy  3660 tccagcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtgga 3719

Db  3900 GCTGGTGACCTTTGACCAGATGGTGAACCTCTCCATCGATGGTGGCAGCCCCATGACCAT 3959
Qy  3720 actacttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcat 3779

Db  3960 GGACAACTTTGGAAAGCACTACACACTCAACAGTGAGGCCCCCCTCTATGTGGGAGGGAT 4019
Qy  3780 cactaacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcat 3839

Db  4020 GCCCGTGGATGTGAACTCAGCTGCCTTCCGCCTGTGGCAGATCCTCAATGGCACCAGCTT 4079
Qy  3840 gccagggaagagtaacgtggcatctctgcgccaggcccctgggcagaacggaaccagctt 3899

Db  4080 CCACGGTTGCATCCGGAATCTATACATCAACAACGAACTGCAGGACTTCACCAAGACACA 4139
Qy  3900 ccacggctgcatccggaacctttacatcaacagtgagctgcaggacttccagaaggtgcc 3959

Db  4140 GATGAAGCCGGGCGTGGTGCCCGGCTGCGAGCCCTGCCGAAAACTCTACTGTCTACATGG 4199
Qy  3960 gatgcaaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatgg 4019

Db  4200 CATTTGCCAGCCCAACGCCACCCCAGGGCCCGTGTGCCACTGCGAGGCTGGCTGGGGGGG 4259
Qy  4020 cacatgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggg 4079

Db  4260 CCTGCACTGTGACCAGCCAGTGGACGGCCCCTGCCATGGCCACAAGTGTGTCCATGGGAA 4319
Qy  4080 gcccctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcac 4139

Db  4320 ATGCGTGCCCCTCGATGCACTTGCCTACAGCTGCCAATGCCAGGATGGATACTCGGGGGC 4379
Qy  4140 ctgcttgcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgt 4199

Db  4380 TCTGTGCAACCAGGTCGGGGCTGTGGCAGAGCCCTGTGGGGGCTTACAGTGCCTGCATGG 4439
Qy  4200 cctctgtgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatgg 4259

Db  4440 TCACTGCCAGGCCTCGGCCACCAGAGGGGCACACTGTGTGTGCAGCCCAGGTTTTTCAGG 4499
Qy  4260 gaagtgcaggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacggg 4319

Db  4500 CGAGCTGTGTGAGCAAGAGTCCGAGTGCCGGGGGGACCCTGTCCGGGACTTTCACCGGGT 4559
Qy  4320 ggacagctgtgatcgagaaatctcttgtcgaggggaaaggataagagattattaccaaaa 4379

Db  4560 CCAGAGGGGCTATGCCATCTGCCAGACCACGCGCCCACTGTCATGGGTGGAATGCCGGGG 4619
Qy  4380 gcagcagggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagagg 4439

Db  4620 CGCGTGCCCGGGCCAGGGCTGCTGCCAGGGCCTGCGGCTGAAGCGGAGGAAGCTCACCTT 4679
Qy  4440 tgggtgtgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactcttt 4499

Db  4680 CGAGTGCAGCGATGGGACCTCGTTTGCTGAGGAGGTGGAGAAGCCCACCAAGTGCGGCTG 4739
Qy  4500 cgaatgcactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559

Db  4740 TGCCCCGTGTGCGT 4753
Qy  4560 tacgaggtgtgtgt 4573
```

RESULT 8
LOCUS AF075240 2553 bp mRNA PRI 09-MAR-1999
DEFINITION Homo sapiens SLIT1 protein (SLIL2) mRNA, partial cds.
ACCESSION AF075240
NID g4377994
VERSION AF075240.1 GI:4377994
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2553)
AUTHORS Holmes,G.P., Negus,K., Burridge,L., Raman,S., Algar,E., Yamada,T. and Little,M.H.
TITLE Distinct but overlapping expression patterns of two vertebrate slit homologs implies functional roles in CNS development and organogenesis
JOURNAL Mech. Dev. 79 (1-2), 57-72 (1998)
REFERENCE 2 (bases 1 to 2553)
AUTHORS Holmes,G.P., Negus,K., Raman,S. and Little,M.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1998) Centre for Molecular and Cellular Biology, University of Queensland, St. Lucia, Brisbane, Queensland 4072, Australia
FEATURES Location/Qualifiers
source 1. .2553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:T65521"
/chromosome="5"
/map="5q35"
/clone="IMAGE Consortium ID 21651"
gene <1. .2553
/note="SLIT1"
/gene="SLIL2"
CDS <1. .2553
/gene="SLIL2"
/note="extracelluar leucine-rich protein; contains leucine-rich and EGF-like repeats; similar to Drosophila neurogenic extracellular slit protein"
/codon_start=1

```
                /product="SLIT1 protein"
                /protein_id="AAD19336.1"
                /db_xref="PID:g4377995"
                /db_xref="GI:4377995"
                /translation="EGAFDGAASVQELMLTGNQLETVHGRVFRGLSGLKTLMLRSNLI
                GCVSNDTFAGLSSVRLLSLYDNRITTITPGAFTTLVSLSTINLLSNPFNCNCHLAWLG
                KWLRKRRIVSGNPRCQKPFFLKEIPIQDVAIQDFTCDGNEESSCQLSPRCPEQCTCME
                TVVRCSNKGLRALPRGMPKDVTELYLEGNHLTAVPRELSALRHLTLIDLSNNSISMLT
                NYTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVLTLHGNDISSVPEGSFNDLTSL
                SHLALGTNPLHCDCSLRWLSEWVKAGYKEPGIARCSSPEPMADRLLLTTPTHRFQCKG
                PVDINIVAKCNACLSSPCKNNGTCTQDPVELYRCACPYSYKGKDCTVPINTCIQNPCQ
                HGGTCHLSDSHKDGFSCSCPLGFEGQRCEINPDDCEDNDCENNATCVDGINNYVCICP
                PNYTGELCDEVIDHCVPELNLCQHEAKCIPLDKGFSCECVPGYSGKLCETDNDDCVAH
                KCRHGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVYDSLSSPPTTVYSVETVNDGQFH
                SVELVTLNQTLNLVVDKGTPKSLGKLQKQPAVGINSPLYLGGIPTSTGLSALRQGTDR
                PLGGFHGCIHEVRINNELQDFKALPPQSLGVSPGCKSCTVCKHGLCRSVEKDSVVCEC
                RPGWTGPLCDQEARDPCLGHRCHHGKCVATGTSYMCKCAEGYGGDLCDNKNDSANACS
                AFKCHHGQCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYASCATAS
                KVPIMECRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS"
misc_feature    10. .69
                /gene="SLIL2"
                /note="Region: LRR3-2"
misc_feature    82. .141
                /gene="SLIL2"
                /note="Region: LRR3-3"
misc_feature    154. .213
                /gene="SLIL2"
                /note="Region: LRR3-4"
misc_feature    226. .261
                /gene="SLIL2"
                /note="Region: LRR3-5"
misc_feature    247. .264
                /gene="SLIL2"
                /note="Region: LRR4-4"
misc_feature    268. .453
                /gene="SLIL2"
                /note="Region: LRR carboxy-flanking region"
misc_feature    454. .549
                /gene="SLIL2"
                /note="Region: LRR amino-flanking region"
misc_feature    550. .582
                /gene="SLIL2"
                /note="Region: LRR4-1"
misc_feature    592. .651
                /gene="SLIL2"
                /note="Region: LRR4-2"
misc_feature    664. .723
                /gene="SLIL2"
                /note="Region: LRR4-3"
misc_feature    808. .843
                /gene="SLIL2"
                /note="Region: LRR4-5"
misc_feature    850. .1023
                /gene="SLIL2"
                /note="Region: LRR carboxy-flanking region"
misc_feature    1033. .1140
                /gene="SLIL2"
                /note="Region: EGF1"
misc_feature    1144. .1263
                /gene="SLIL2"
                /note="Region: EGF2"
misc_feature    1267. .1377
                /gene="SLIL2"
                /note="Region: EGF3"
misc_feature    1381. .1497
                /gene="SLIL2"
                /note="Region: EGF4"
misc_feature    1501. .1611
                /gene="SLIL2"
                /note="Region: EGF5"
misc_feature    1966. .2076
                /gene="SLIL2"
                /note="Region: EGF6A"
misc_feature    2083. .2190
                /gene="SLIL2"
                /note="Region: EGF7"
misc_feature    2203. .2313
                /gene="SLIL2"
                /note="Region: EGF8"
misc_feature    2413. .2550
                /gene="SLIL2"
                /note="Region: CTC Knot"
BASE COUNT      561 a    798 c    719 g    475 t
ORIGIN

  Query Match          10.6%;  Score 505;  DB 31;  Length 2553;
  Best Local Similarity  65.9%;  Pred. No. 0.00e+00;
  Matches  1076;  Conservative    0;  Mismatches 553;  Indels   3;  Gaps   1;

Db      1 GAGGGAGCTTTCGATGGAGCAGCCAGCGTGCAGGAGCTGATGCTGACAGGGAACCAGCTG 60
          |||||||| || || |||||| |  | ||  | ||  |  | | || ||  | || |   ||
Qy   1711 gagggagcatttgaaggagcatctggtgtaaatgaaatacttcttacgagtaatcgtttg 1770

Db     61 GAGACCGTGCACGGGCGCGTGTTCCGTGGCCTCAGTGGCCTCAAAACCTTGATGCTGAGG 120
          || |  |||||        |||||  ||  |     |||||||||| |||||| ||||
Qy   1771 gaaaatgtgcagcataagatgttcaagggattggaaagcctcaaaactttgatgttgaga 1830

Db    121 AGTAACTTGATCGGCTGTGTGAGTAATGACACCTTTGCCGGCCTGAGTTCGGTGAGACTG 180
          || ||    ||  ||||||| | |||||||  ||    || || ||||| ||| |  ||
Qy   1831 agcaatcgaataacctgtgtggggaatgacagtttcataggactcagttctgtgcgtttg 1890

Db    181 CTGTCCCTCTATGACAATCGGATCACCACCATCACCCCTGGGGCCTTCACCACGCTTGTC 240
          || ||  | ||||| ||||  || || ||  |  | || |||||  ||    || ||
Qy   1891 ctttctttgtatgataatcaaattactacagttgcaccaggggcatttgatactctccat 1950

Db    241 TCCCTGTCCACCATAAACCTCCTGTCCAACCCCTTCAACTGCAACTGCCACCTGGCCTGG 300
          ||  | || ||  |||||||| || |||| || || ||||| |||||| ||||||| |||
Qy   1951 tctttatctactctaaacctcttggccaatccttttaactgtaactgctacctggcttgg 2010

Db    301 CTCGGCAAGTGGTTGAGGAAGAGGCGGATCGTCAGTGGGAACCCTAGGTGCCAGAAGCCA 360
           | ||  ||||| |||| |||| | | || ||||  || || ||||| || || || |||
Qy   2011 ttgggagagtggctgagaaagaagagaattgtcacgggaaatcctagatgtcaaaaacca 2070

Db    361 TTTTTCCTCAAGGAGATTCCCATCCAGGATGTGGCCATCCAGGACTTCACCTGTGATGGC 420
          |  ||||| || || || ||||||||||||||||||||| |||||||||| ||||||| |
Qy   2071 tacttcctgaaagaaatacccatccaggatgtggccattcaggacttcacttgtgatgac 2130

Db    421 ---AACGAGGAGAGTAGCTGCCAGCTGAGCCCGCGCTGCCCGGAGCAGTGCACCTGTATG 477
              || || || | ||| |||   |      | |||||  ||     | || || ||  ||
Qy   2131 ggaaatgatgacaatagttgctccccactttctcgctgtcctactgaatgtacttgcttg 2190

Db    478 GAGACAGTGGTGCGATGCAGCAACAAGGGGCTCCGCGCCCTCCCCAGAGGCATGCCCAAG 537
          || ||||| || ||||| ||||||||||  |    | | | || | ||| || || |
Qy   2191 gatacagtcgtccgatgtagcaacaagggtttgaaggtcttgccgaaaggtattccaaga 2250

Db    538 GATGTGACCGAGCTGTACCTGGAAGGAAACCACCTAACAGCCGTGCCCAGAGAGCTGTCC 597
          ||||| || ||| |||| ||||| ||||||||  | |||   || ||||  || || |||
Qy   2251 gatgtcacagagttgtatctggatggaaaccaatttacactggttcccaaggaactctcc 2310

Db    598 GCCCTCCGACACCTGACGCTTATTGACCTGAGCAACAACAGCATCAGCATGCTGACCAAT 657
            |  | |||  | || ||||| ||| | || |||||||| || |||| |||  | |||
Qy   2311 aactacaaacatttaacacttatagacttaagtaacaacagaataagcacgctttctaat 2370

Db    658 TACACCTTCAGTAACATGTCTCACCTCTCCACTCTGATCCTGAGCTACAACCGGCTGAGG 717
          | | |||||| |||||| | || |||  |||  | || || || |||||||| |||||
Qy   2371 cagagcttcagcaacatgacccagctcctcaccttaattcttagttacaaccgtctgaga 2430

Db    718 TGCATCCCCGTCCACGCCTTCAACGGGCTGCGGTCCCTGCGAGTGCTAACCCTCCATGGC 777
          || || ||    | | |||| | || |  ||| || ||| | ||  | || |||||
Qy   2431 tgtattcctcctcgcaccttgatggattaaagtctcttcgattactttctctacatgga 2490

Db    778 AATGACATTTCCAGCGTTCCTGAAGGCTCCTTCAACGACCTCACATCTCTTTCCCATCTG 837
          ||||||||||    || ||||||||  | ||||| || ||  |  |  | || |||||
Qy   2491 aatgacatttctgttgtgcctgaaggtgctttcaatgatctttctgcattatcacatcta 2550

Db    838 GCGCTGGGAACCAACCCACTCCACTGTGACTGCAGTCTTCGGTGGCTGTCGGAGTGGGTG 897
```

```
Qy   2551 gcaattggagccaaccctctttactgtgattgtaacatgcagtggttatccgactgggtg 2610

Db    898 AAGGCGGGGTACAAGGAGCCTGGCATCGCCCGCTGCAGTAGCCCTGAGCCCATGGCTGAC 957
Qy   2611 aagtcggaatataaggagcctggaattgctcgttgtgctggtcctggagaaatggcagat 2670

Db    958 AGGCTCCTGCTCACCACCCCAACCCACCGCTTCCAGTGCAAAGGGCCAGTGGACATCAAC 1017
Qy   2671 aaacttttactcacaactccctcccaaaaaatttacctgtcaaggtcctgtggatgtcaat 2730

Db   1018 ATTGTGGCCAAATGCAATGCCTGCCTCTCCAGCCCGTGCAAGAATAACGGGACATGCACC 1077
Qy   2731 attctagctaagtgtaacccctgcctatcaaatccgtgtaaaaatgatggcacatgtaat 2790

Db   1078 CAGGACCCTGTGGAGCTGTACCGCTGTGCCTGCCCCTACAGCTACAAGGGCAAGGACTGC 1137
Qy   2791 agtgatccagttgacttttaccgatgcacctgtccatatggtttcaaggggcaggactgt 2850

Db   1138 ACTGTGCCCATCAACACCTGCATCCAGAACCCCTGTCAGCATGGAGGCACCTGCCACCTG 1197
Qy   2851 gatgtcccaattcatgcctgcatcagtaacccatgtaaacatggaggaacttgccactta 2910

Db   1198 AGTGACAGCCACAAGGATGGGTTCAGCTGCTCCTGCCCCTCTGGGCTTTGAGGGGCAGCGG 1257
Qy   2911 aaggaaggagaagaagatggattctggtgtatttgtgctgatggatttgaaggagaaaat 2970

Db   1258 TGTGAGATCAACCCAGATGACTGTGAGGACAACGACTGCGAAAACAATGCCACCTGCGTG 1317
Qy   2971 tgtgaagtcaacgttgatgattgtgaagataatgactgtgaaaataattctacatgtgtc 3030

Db   1318 GACGGGATCAACAACTACGTGTGTATCTGTCCGCCTAACTACACAGGTGAGCTATGCGAC 1377
Qy   3031 gatggcattaataactacacatgcctttgcccacctgagtatacaggtgagttgtgtgag 3090

Db   1378 GAGGTGATTGACCACTGTGTGCCTGAGCTGAACCTCTGTCAGCATGAGGCCAAGTGCATC 1437
Qy   3091 gagaagctggacttctgtgcccaggacctgaacccctgccagcacgattcaaagtgcatc 3150

Db   1438 CCCCTGGACAAAGGATTCAGCTGCGAGTGTGTCCCTGGCTACAGCGGGAAGCTCTGTGAG 1497
Qy   3151 ctaactccaaagggattcaaatgtgactgcacaccagggtacgtaggtgaacactgcgac 3210

Db   1498 ACAGACAATGATGACTGTGTGGCCCACAAGTGCCGCCACGGGGCCCAGTGCGTGGACACA 1557
Qy   3211 atcgattttgacgactgccaagacaacaagtgtaaaaacggagcccactgcacagatgca 3270

Db   1558 ATCAATGGCTACACATGCACCTGCCCCCAGGGCTTCAGTGGACCCTTCTGTGAACACCCC 1617
Qy   3271 gtgaacggctatacgtgcatatgccccgaaggttacagtggcttgttctgtgagtttttct 3330

Db   1618 CCACCCATGGTC 1629
Qy   3331 ccacccatggtc 3342
```

```
RESULT    9
LOCUS       AB011538     6921 bp    mRNA           PRI       22-AUG-1998
DEFINITION  Homo sapiens mRNA for MEGF5, partial cds.
ACCESSION   AB011538
NID         g3449301
VERSION     AB011538.1  GI:3449301
KEYWORDS    MEGF5.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
            plus clone:HG2635.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6921)
  AUTHORS   Nakayama,M., Nakajima,D. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1998) to the DDBJ/EMBL/GenBank databases. Manabu
            Nakayama, Kazusa DNA Research Institute, Laboratory of DNA
            technology; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:nmanabu@kazusa.or.jp, Tel:+81-438-52-3915,
            Fax:+81-438-52-3914)
REFERENCE   2  (sites)
  AUTHORS   Nakayama,M., Nakajima,D., Nagase,T., Nomura,N., Seki,N. and
            Ohara,O.
  TITLE     Identification of high-molecular-weight proteins with multiple
            EGF-like motifs by motif-trap screening
  JOURNAL   Genomics 51 (1), 27-34 (1998)
  MEDLINE   98360089
FEATURES             Location/Qualifiers
     source          1. .6921
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="HG2635"
                     /clone_lib="pBluescriptII SK plus"
                     /map="5q35"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .2221
                     /gene="MEGF5"
     CDS             <1. .2221
                     /gene="MEGF5"
                     /note="human homologue of Drosophila slit protein; human
                     slit2"
                     /codon_start=2
                     /product="MEGF5"
                     /protein_id="BAA32466.1"
                     /db_xref="PID:d1033429"
                     /db_xref="PID:g3449302"
                     /db_xref="GI:3449302"
                     /translation="NSISMLTNYTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVL
                     TLHGNDISSVPEGSFNDLTSLSHLALGTNPLHCDCSLRWLSEWVKAGYKEPGIARCSS
                     PEPMADRLLLTTPTHRFQCKGPVDINIVAKCNACLSSPCKNNGTCTQDPVELYRCACP
                     YSYKGKDCTVPINTCIQNPCQHGGTCHLSDSHKDGFSCSCPLGFEGQRCEINPDDCED
                     NDCENNATCVDGINNYVCICPPNYTGELCDEVIDHCVPELNLCQHEAKCIPLDKGFSC
                     ECVPGYSGKLCETDNDDCVAHKCRHGAQCVDTINGYTCTCPQGFSGPFCEHPPPMVLL
                     QTSPCDQYECQNGAQCIVVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKV
                     RPQANISLQVATDKDNGILLYKGDNDPLALELYQGHVRLVYDSLSSPPTTVYSVETVN
                     DGQFHSVELVTLNQTLNLVVDKGTPKSLGKLQKQPAVGINSPLYLGGIPTSTGLSALR
                     QGTDRPLGGFHGCIHEVRINNELQDFKALPPQSLGVSPGCKSCTVCKHGLCRSVEKDS
                     VVCECRPGWTGPLCDQEARDPCLGHRCHHGKCVATGTSYMCKCAEGYGGDLCDNKNDS
                     ANACSAFKCHHGQCHISDQGEPYCLCQPGFSGEHCQQENPCLGQVVREVIRRQKGYAS
                     CATASKVPIMECRGGCGPQCCQPTRSKRRKYVFQCTDGSSFVEEVERHLECGCLACS"
     polyA_site      6921
                     /note="19 a nucleotides"
BASE COUNT     1820 a   1813 c   1708 g   1580 t
ORIGIN
```

```
  Query Match           9.9%;  Score 469;  DB 29;  Length 6921;
  Best Local Similarity  61.6%;  Pred. No. 0.00e+00;
  Matches  1364;  Conservative    0;  Mismatches 841;  Indels   9;  Gaps   7;

Db      1 CAACAGCATCAGCATGCTGACCAATTACACCTTCAGTAACATGTCTCACCTCTCCACTCT 60
Qy   2346 caacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcctcaccttt 2405

Db     61 GATCCTGAGCTACAACCGGCTGAGGTGCATCCCCGTCCACGCCTTCAACGGGCTGCGGTC 120
Qy   2406 aattcttagttacaaccgtctgagatgtattcctcctcgcacctttgatggattaaagtc 2465

Db    121 CCTGCGAGTGCTAACCCTCCATGGCAATGACATTTCCAGCGTTCCTGAAGGCTCCTTCAA 180
Qy   2466 tcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgctttcaa 2525

Db    181 CGACCTCACATCTCTTTCCCATCTGGCGCTGGGAACCAACCCACTCCACTGTGACTGCAG 240
Qy   2526 tgatctttctgcattatcacatctagcaattggagccaaccctctttactgtgattgtaa 2585

Db    241 TCTTCGGTGGCTGTCGGAGTGGGTGAAGGCGGGGTACAAGGAGCCTGGCATCGCCCGCTG 300
Qy   2586 catgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgctcgttg 2645
```

```
Db    301 CAGTAGCCCTGAGCCCATGGCTGACAGGCTCCTGCTCACCACCCCAACCCACCGCTTCCA 360
Qy   2646 tgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaaatttac 2705

Db    361 GTGCAAAGGGCCAGTGGACATCAACATTGTGGCCAAATGCAATGCCTGCCTCTCCAGCCC 420
Qy   2706 ctgtcaaggtcctgtggatgtcaatattctagctaagtgtaacccctgcctatcaaatcc 2765

Db    421 GTGCAAGAATAACGGGACATGCACCCAGGACCCTGTGGAGCTGTACCGCTGTGCCTGCCC 480
Qy   2766 gtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcacctgtcc 2825

Db    481 CTACAGCTACAAGGGCAAGGACTGCACTGTGCCCATCAACACCTGCATCCAGAACCCCTG 540
Qy   2826 atatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaacccatg 2885

Db    541 TCAGCATGGAGGCACCTGCCACCTGAGTGACAGCCACAAGGATGGGTTCAGCTGCTCCTG 600
Qy   2886 taaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtgtatttg 2945

Db    601 CCCTCTGGGCTTTGAGGGGCAGCGGTGTGAGATCAACCCAGATGACTGTGAGGACAACGA 660
Qy   2946 tgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaagataatga 3005

Db    661 CTGCGAAAACAATGCCACCTGCGTGGACGGGATCAACAACTACGTGTGTATCTGTCCGCC 720
Qy   3006 ctgtgaaaataattctacatgtgtcgatggcattaataactacacatgcctttgcccacc 3065

Db    721 TAACTACACAGGTGAGCTATGCGACGAGGTGATTGACCACTGTGTGCCTGAGCTGAACCT 780
Qy   3066 tgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacctgaaccc 3125

Db    781 CTGTCAGCATGAGGCCAAGTGCATCCCCCTGGACAAAGGATTCAGCTGCGAGTGTGTCCC 840
Qy   3126 ctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactgcacacc 3185

Db    841 TGGCTACAGCGGGAAGCTCTGTGAGACAGACAATGATGACTGTGTGGCCCACAAGTGCCG 900
Qy   3186 agggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaagtgtaa 3245

Db    901 CCACGGGGCCCAGTGCGTGGACACAATCAATGGCTACACATGCACCTGCCCCCAGGGCTT 960
Qy   3246 aaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccgaaggtta 3305

Db    961 CAGTGGACCCTTCTGTGAACACCCCCCACCCATGGTCCTACTGCAGACCAGCCCATGCGA 1020
Qy   3306 cagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagcccctgtga 3365

Db   1021 CCAGTACGAGTGCCAGAACGGGGCCCAGTGCATCGTGGTGCAGCAGGAGCCCACCTGCCG 1080
Qy   3366 taattttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaatatgtca 3425

Db   1081 CTGCCCACCAGGCTTCGCCGGCCCCAGATGCGAGAAGCTCATCACTGTCAACTTCGTGGG 1140
Qy   3426 gtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaattttataaa 3485

Db   1141 CAAAGACTCCTACGTGGAACTGGCCTCCGCCAAGGTCCGACCCCAGGCCAACATCTCCCT 1200
Qy   3486 caaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacataacact 3545

Db   1201 GCAGGTGGCCACTGACAAGGACAACGGCATCCTTCTCTACAAAGGAGACAATGACCCCCT 1260
Qy   3546 tcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaagaccatat 3605

Db   1261 GGCACTGGAGCTGTACCAGGGCCACGTGCGGCTGGTCTATGACAGCCTGAGTTCCCCTCC 1320
Qy   3606 cgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctcatccagc 3665

Db   1321 AACCACAGTGTACAGTGTGGAGACAGTGAATGATGGGCAGTTTCACAGTGTGGAGCTGGT 1380
Qy   3666 ttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtggaactact 3725

Db   1381 GACGCTAAACCAGACCCTGAACCTAGTAGTGGACAAAGGAACTCCAAAGAGCCTGGGGAA 1440
Qy   3726 tgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcatcactaa 3785

Db   1441 GCTCCAGAAGCAGCCAGCAGTGGGCATCAACAGCCCCCTCTACCTTGGAGGCATCCCCAC 1500
Qy   3786 cttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcatgccagg 3845

Db   1501 CTCCACCGGCCTCTCCGCCTTGCGCCAGGGCACGGACCGGCCTCTAGGCGGCTTCCACGG 1560
Qy   3846 gaagagtaacgtggcatctctgcgccaggcccctgggcagaacggaaccagcttccacgg 3905

Db   1561 ATGCATCCATGAGGTGCGCATCAACAACGAGCTGCAGGACTTCAAGGCCCTCCCACCACA 1620
Qy   3906 ctgcatccggaaccttacatcaacagtgagctgcaggacttccagaaggtgccgatgca 3965

Db   1621 GTCCCTGGGGGTGTCAC-CAGGCTGCAA-GTC-CTGCACCGTGTGCAAGCACGGCCTGTG 1677
Qy   3966 aacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatggcacatg 4025

Db   1678 CCGCTCCGTGGAGAAGGACAGCGTGGTGTGCGAGTGCCGCCCAGGCTGGACCGGCCCACT 1737
Qy   4026 ccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggggcccct 4085

Db   1738 CTGCGACCAGGAGGCCCGGGACCCCTGCCTCGGCCACAGATGCCACCATGGAAAATGTGT 1797
Qy   4086 ctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcacctgctt 4145

Db   1798 GGCAA-C--TGGGACCTCATACATGTGCAAGTGTGCCGAGGGCTATGGAGGGGACTTGTG 1854
Qy   4146 gcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgtcctctg 4205

Db   1855 TGACAACAAGAATGACTCTGCCAATGCCTGCTCAGCCTTCAAGTGTCACCATGGGCAGTG 1914
Qy   4206 tgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatgggaagtg 4265

Db   1915 CCACATCTCAGACCAAGGGGAGCCCTACTGCCTGTGCCAGCCCGGCTTTAGCGGCGAGCA 1974
Qy   4266 caggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacgggggacag 4325

Db   1975 CTGCCAACAAGAGAATCCGTGCCTGGGACAAGTAGTCCGAGAGGTGATCCGCCGCCAGAA 2034
Qy   4326 ctgtgatcgagaaaatctcttgtcgaggggaaaggataagagattattaccaaaagcagca 4385

Db   2035 AGGTTATGCATCATGTGCCACAGCCTCCAAGGTGCCCATCATGGAATGTCGTGGGGGCTG 2094
Qy   4386 gggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagaggtgggtg 4445

Db   2095 TG--GGCCC-CAGTGCTGCCAGCCCACCCGCAGCAAGCGGCGGAAATACGTCTTCCAGTG 2151
Qy   4446 tgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactctttcgaatg 4505

Db   2152 CACGGACGGCTCCTCGTTTGTAGAAGAGGTGGAGAGACACTTAGAGTGCGGCTG 2205
Qy   4506 cactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559
```

```
RESULT   10
LOCUS       AB017170      591 bp    mRNA            ROD       06-FEB-1999
DEFINITION  Rattus norvegicus mRNA for Slit-1 protein, partial cds.
ACCESSION   AB017170
NID         g4049590
VERSION     AB017170.1  GI:4049590
KEYWORDS    slit-1; Slit-1 protein.
SOURCE      Rattus norvegicus tissue_lib:brain cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Itoh,A. and Sakano,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1998) to the DDBJ/EMBL/GenBank databases. Akira
            Itoh, Asahi Chemical Industry co.,ltd., Life Science Fundamental
```

```
          Research Laboratory; 2-1, Samejima, Fuji, Shizuoka 416-8501, Japan
          (E-mail:a8611483@ut.asahi-kasei.co.jp, Tel:+81-545-62-3231,
          Fax:+81-545-62-3249)
REFERENCE 2  (sites)
  AUTHORS   Itoh,A., Miyabayashi,T., Ohno,M. and Sakano,S.
  TITLE     Cloning and expressions of three mammalian homologues of Drosophila
            slit suggest possible roles for Slit in the formation and
            maintenance of the nervous system
  JOURNAL   Brain Res. Mol. Brain Res. 62 (2), 175-186 (1998)
  MEDLINE   99033071
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_lib="brain"
     gene            2. .589
                     /gene="slit-1"
     CDS             <2. .>589
                     /gene="slit-1"
                     /codon_start=1
                     /product="Slit-1 protein"
                     /protein_id="BAA35187.1"
                     /db_xref="PID:d1036173"
                     /db_xref="PID:g4049591"
                     /db_xref="GI:4049591"
                     /translation="IAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVTFD
                     QMVNLSIDGGSPMTMDNFGKHYTLNSEAPLYVGGMPVDVNSAAFRLWQILNGTSFHGC
                     IRNLYINNELQDFTKTQMKPGVVPGCEPCRKLYCLHGICQPNATPGPVCHCEAGWGGL
                     HCDQPVDGPCHGHKCVHGKCVPLDALAYSCQCQDGY"
BASE COUNT      129 a    185 c    161 g    116 t
ORIGIN

  Query Match           3.8%;  Score 182;  DB 32;  Length 591;
  Best Local Similarity  65.6%;  Pred. No. 1.15e-110;
  Matches   383;  Conservative    0;  Mismatches 201;  Indels   0;  Gaps   0;

Db      2 ATTGCAGTTGAGCTGTACCAGGGCCATGTCCGTGTTAGCTACGACCCAGGCAGCTACCCC 61
          || || || || || || | ||| | ||| |||| ||||| ||| | |||    | ||
Qy   3604 atcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctcatcca 3663

Db     62 AGCTCTGCTATCTACAGTGCTGAAACAATCAACGATGGGCAGTTCCACACAGTTGAGCTG 121
           ||||| || ||||||| || |||||||| |||||  | |||||||  || || ||
Qy   3664 gcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtggaacta 3723

Db    122 GTGACCTTTGACCAGATGGTGAACCTCTCCATCGATGGTGGCAGCCCCATGACCATGGAC 181
          |  |||| || ||||   |     | ||| | |||||||| | |||||  | |||
Qy   3724 cttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcatcact 3783

Db    182 AACTTTGGAAAGCACTACACACTCAACAGTGAGGCCCCCCTCTATGTGGGAGGGATGCCC 241
          |||||   |||||| | ||| || ||   |||  | || |||||||| ||||| |||||
Qy   3784 aacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcatgcca 3843

Db    242 GTGGATGTGAACTCAGCTGCCTTCCGCCTGTGGCAGATCCTCAATGGCACCAGCTTCCAC 301
          | | |   |||  ||  |  | |||| |   |     |  || || ||||||||||||
Qy   3844 gggaagagtaacgtggcatctctgcgccaggccccctgggcagaacggaaccagcttccac 3903

Db    302 GGTTGCATCCGGAATCTATACATCAACAACGAACTGCAGGACTTCACCAAGACACAGATG 361
          || |||||||||||| || |||||||||| || ||||||||||||   |||  | ||||
Qy   3904 ggctgcatccggaacctttacatcaacagtgagctgcaggacttccagaaggtgccgatg 3963

Db    362 AAGCCGGGCGTGGTGCCCGGCTGCGAGCCCTGCCGAAAACTCTACTGTCTACATGGCATT 421
          |  | ||| |  |||| ||||| ||||| ||||  ||       |||  |||||||
Qy   3964 caaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatggcaca 4023

Db    422 TGCCAGCCCAACGCCACCCCAGGGCCCGTGTGCCACTGCGAGGCTGGCTGGGGGGGCCTG 481
          |||||||||| |  |    ||||   |  ||| | ||| |||  || |||  ||| |
Qy   4024 tgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggggccc 4083

Db    482 CACTGTGACCAGCCAGTGGACGGCCCCTGCCATGGCCACAAGTGTGTCCATGGGAAATGC 541
          | |||||||||  |      | | ||| |||| |||  | || || || ||||| |  |||
Qy   4084 ctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcacctgc 4143

Db    542 GTGCCCCTCGATGCACTTGCCTACAGCTGCCAGTGCCAGGATGG 585
```

```
          ||||| || ||||  |  ||||||||||  |||||  ||| ||
Qy   4144 ttgcccatcaatgcgttctcctacagctgtaagtgcttggaggg 4187


RESULT   11
LOCUS       DMSLIT      5401 bp    mRNA            INV       14-FEB-1991
DEFINITION  Drosophila mRNA for slit protein.
ACCESSION   X53959
NID         g8614
VERSION     X53959.1  GI:8614
KEYWORDS    extracellular protein; morphogenesis protein; slit gene.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5401)
  AUTHORS   Rothberg,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1990) J.M. Rothberg, YALE UNIVERSITY, DEPT. OF
            BIOLOGY, P.O.BOX 6666, NEW HAVEN, CT 06520, USA
REFERENCE   2  (bases 1 to 5401)
  AUTHORS   Rothberg,J.M., Jacobs,J.R., Goodman,C.S. and Artavanis-Tsakonas,S.
  TITLE     slit: an extracellular protein necessary for development of midline
            glia and commissural axon pathways contains both EGF and LRR
            domains
  JOURNAL   Genes Dev. 4 (12A), 2169-2187 (1990)
  MEDLINE   91099665
FEATURES             Location/Qualifiers
     source          1. .5401
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /dev_stage="embryo"
     mRNA            1. .5401
                     /gene="sli"
                     /db_xref="FlyBase:FBgn0003425"
                     /evidence=experimental
     gene            1. .5401
                     /note="slit"
                     /gene="sli"
                     /allele=""
                     /db_xref="FlyBase:FBgn0003425"
     sig_peptide     251. .358
                     /gene="sli"
                     /db_xref="FlyBase:FBgn0003425"
                     /product="slit protein"
     CDS             251. .4693
                     /gene="sli"
                     /codon_start=1
                     /product="slit protein"
                     /protein_id="CAA37910.1"
                     /db_xref="PID:g8615"
                     /db_xref="GI:8615"
                     /db_xref="FLYBASE:FBgn0003425"
                     /db_xref="SWISS-PROT:P24014"
                     /translation="MAAPSRTTLMPPPFRLQLRLLILPILLLLRHDAVHAEPYSGGFG
                     SSAVSSGGLGSVGIHIPGGGVGVITEARCPRVCSCTGLNVDCSHRGLTSVPRKISADV
                     ERLELQGNNLTVIYETDFQRLTKLRMLQLTDNQIHTIERNSFQDLVSLERLDISNNVI
                     TTVGRRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVELEILTLNNNNLTSLPHNIFGG
                     LGRLRALRLSDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLKGQNVADLHDQEF
                     KCSGLTEHAPMECGAENSCPHPCRCADGIVDCREKSLTSVPVTLPDDTTDVRLEQNFI
                     TELPPKSFSSFRRLRRIDLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKG
                     LGSLRLLLLNANEISCIRKDAFRDLHSLSLLSLYDNNIQSLANGTFDAMKSMKTVHLA
                     KNPFICDCNLRWLADYLHKNPIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMK
                     LSGECRMDSDCPAMCHCEGTTVDCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGL
                     FGRLPHLVKLELKRNQLTGIEPNAFEGASHIQELQLGENKIKEISNKMFLGLHQLKTL
                     NLYDNQISCVMPGSFEHLNSLTSLNLASNPFNCNCHLAWFAECVRKKSLNGGAARCGA
                     PSKVRDVQIKDLPHSEFKCSSENSEGCLGDGYCPPSCTCTGTVVACSRNQLKEIPRGI
                     PAETSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFANLTKLSTLIIS
                     YNKLQCLQRHALSGLNNLRVVSLHGNRISMLPEGSFEDLKSLTHIALGSNPLYCDCGL
                     KWFSDWIKLDYVEPGIARCAEPEQMKDKLILSTPSSSFVCRGRVRNDILAKCNACFEQ
                     PCQNQAQCVALPQREYQCLCQPGYHGKHCEFMIDACYGNPCRNNATCTVLEEGRFSCQ
                     CAPGYTGARCETNIDDCLGEIKCQNNATCIDGVESYKCECQPGFSGEFCDTKIQFCSP
```

```
                EFNPCANGAKCMDHFTHYSCDCQAGFHGTNCTDNIDDCQNHMCQNGGTCVDGINDYQC
                RCPDDYTGKYCEGHNMISMMYPQTSPCQNHECKHGVCFQPNAQGSDYLCRCHPGYTGK
                WCEYLTSISFVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELFN
                GRIRVSYDVGNHPVSTMYSFEMVADGKYHAVELLAIKKNFTLRVDRGLARSIINEGSN
                DYLKLTTPMFLGGLPVDPAQQAYKNWQIRNLTSFKGCMKEVWINHKLVDFGNAQRQQK
                ITPGCALLEGEQQEEEDDEQDFMDETPHIKEEPVDPCLENKCRRGSRCVPNSNARDGY
                QCKCKHGQRGRYCDQGEGSTEPPTVTAASTCRKEQVREYYTENDCRSRQPLKYAKCVG
                GCGNQCCAAKIVRRRKVRMVCSNNRKYIKNLDIVRKCGCTKKCY"
    mat_peptide 359. .4690
                /gene="sli"
                /db_xref="FlyBase:FBgn0003425"
                /product="slit protein"
    misc_feature 4495. .4527
                /gene="sli"
                /note="alternatively spliced segment"
                /db_xref="FlyBase:FBgn0003425"
BASE COUNT     1313 a   1504 c   1528 g   1056 t
ORIGIN

  Query Match           2.2%;  Score 106;  DB 21;  Length 5401;
  Best Local Similarity  59.5%;  Pred. No. 2.11e-52;
  Matches   331;  Conservative    0;  Mismatches 225;  Indels   0;  Gaps   0;

Db   1132 CTGTCCGCACCCATGTCGCTGTGCGGACGGGATCGTCGATTGCCGTGAGAAGAGTCTGAC 1191
          ||| ||   | | |||  ||||    ||   ||||| || || |||| |||  |||| ||
Qy    828 ctgccctgccgcctgtacctgtagcaacaatatcgtagactgtcgtgggaaaggtctcac 887

Db   1192 CAGCGTGCCCGTCACCTTGCCCGACGACACCACCGACGTTCGCCTCGAGCAAAATTTCAT 1251
             | |||   |   | || ||   || ||| ||  | ||  | || || ||    ||
Qy    888 tgagatccccacaaatcttccagagaccatcacagaaatacgtttggaacagaacacaat 947

Db   1252 TACGGAACTGCCGCCGAAATCGTTCTCCAGCTTTCGACGACTGCGACGCATCGACCTGTC 1311
          |  |   | || ||   | | |||||    | |  |   ||  |||| || ||||||
Qy    948 caaagtcatccctcctggagctttctcaccatataaaaagcttagacgaattgacctgag 1007

Db   1312 CAACAACAACATATCCCGGATTGCCCACGATGCACTAAGCGGCCTAAAGCAGTTAACCAC 1371
          ||| ||  | || ||     |||| |  |||||  |    || |||       | |   |
Qy   1008 caataatcagatctctgaacttgcaccagatgctttccaaggactacgctctctgaattc 1067

Db   1372 TCTCGTGCTGTACGGCAATAAAATAAAGGATTTACCCTCGGGCGTGTTCAAAGGACTCGG 1431
          || || || || || |||||||| |  ||  | |||    |  | ||  |||||||
Qy   1068 acttgtcctctatggaaataaaatcacagaactccccaaaagtttatttgaaggactgtt 1127

Db   1432 CTCGCTCAGGCTGCTGCTGCTGAACGCCAACGAGATCTCGTGCATACGCAAGGATGCCTT 1491
          ||  |   ||| ||  |  |||| |||||| ||||    ||| | ||    |||||  ||
Qy   1128 ttccttacagctcctattattgaatgccaacaagataaactgccttcgggtagatgcttt 1187

Db   1492 TCGCGACCTGCACAGTTTGAGCCTGCTCTCCCTGTACGACAACAACATCCAGTCGCTGGC 1551
          ||  || || ||||  |||| ||| ||||||||  || ||||||||  | ||| |  | ||
Qy   1188 tcaggatctccacaacttgaaccttctctccctatatgacaacaagcttcagaccatcgc 1247

Db   1552 TAATGGCACATTCGACGCCATGAAGAGCATGAAAACGGTACATCTGGCCAAGAATCCTTT 1611
          || || || ||     |  |   |  |||  ||||  | ||| ||||| |||| || ||
Qy   1248 caaggggaccttttcacctcttcgggccattcaaactatgcatttggcccagaacccctt 1307

Db   1612 CATCTGCGACTGCAATCTGCGCTGGCTGGCCGACTATTTGCACAAAAATCCCATAGAGAC 1671
          || || ||||| ||||    ||||| || || ||| | || |  || || || |||||
Qy   1308 tatttgtgactgccatctcaagtggctagcggattatctccatacccaacccgattgagac 1367

Db   1672 GAGTGGAGCCCGCTGC 1687
          ||||| ||||| |||
Qy   1368 cagtggtgcccgttgc 1383


RESULT  12
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match           1.7%;  Score 80;  DB 25;  Length 7218;
  Best Local Similarity  2.4%;  Pred. No. 1.95e-33;
  Matches    9;  Conservative  221;  Mismatches 150;  Indels   0;  Gaps   0;

Db   1052 GAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1111
          ||| | | ||  ||  : : :::: : :  :: : :::: :: :::::::  :::: :::
Cp   4586 gagtgtgtttaggacacacacctcgtacagccgcacttcaccactttctcaacctcgtcc 4527

Db   1112 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1171
           :          :: ::  : : :::       : ::::: :: ::: ::::::  :  :
Cp   4526 acaaaggaggagccgtcagtgcattcgaaagagtatttccgccgcttgctcctcagcggt 4467

Db   1172 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1231
          :: :  : :: ::::::: : : ::: ::::: : ::::  ::    : :::::::  ::
Cp   4466 ccacagcactgccctcctgcacacccacctctgcactctaatcgggacaccttcttggtt 4407

Db   1232 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1291
          :::  :   :  : :  :::: :: :::::  :  :  :::::: ::::::::::::
Cp   4406 gtttggcaagcagcatagccctgctgcttttggtaataatctcttatcctttcccctcga 4347

Db   1292 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1351
          :      :::::  :: :  :: :::::: : : ::: :: :: : ::: :  :    :
Cp   4346 caagagatttctcgatcacagctgtccccccgtgtatccactgctgcattcacagtagggc 4287

Db   1352 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1411
          : ::::   :::     ::: : :::::: : ::: : :::  :: :::  : :   ::
Cp   4286 tgccccagacctgaaagcctgcacttcccatgcttgcacttgatcgcctggcatgggtta 4227

Db   1412 YYYYYYYYYYYYYYYYYY 1431
          :  :::::::::::: ::
Cp   4226 aacagatcctcctcttcatc 4207


RESULT  13
LOCUS       G41330       297 bp    DNA             STS       26-AUG-1998
DEFINITION  Z1154 Zebrafish AB Danio rerio STS genomic, sequence tagged site.
ACCESSION   G41330
NID         g3461906
VERSION     G41330.1  GI:3461906
KEYWORDS    STS.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 297)
  AUTHORS   Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.
            and Fishman,M.C.
  TITLE     A genetic linkage map of the zebrafish with 2000 microsatellite
            markers
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mark C. Fishman
            Cardiovascular Research Center
            Massachusetts General Hospital
            Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
            Fax: 6177265806
            Email: fishman@mgh.cvrc.harvard.edu
            http://zebrafish.mgh.harvard.edu
            Primer A: TCATGATTGTTTGGAATGTAATAGTG
            Primer B: TTGAGCGGTAGTCTTCTACGC
            STS size: 130
```

```
          PCR Profile:
              Presoak:        94 degrees C for 5.0 minutes
              Denaturation:   94 degrees C for 1.0 minute
              Annealing:      58 degrees C for 1.0 minute
              Polymerization: 72 degrees C for 1.5 minute
              PCR Cycles:     27
              Thermal Cycler: MJ Research PTC-100
          Protocol:
              Template:       10 ng
              Primer:         each 375 nM
              dNTPs:          each 200 uM
              Taq Polymerase: 0.034 units/ul
              Total Vol:      10 ul

          Buffer:
                          MgCL2:          1.5 mM
                          KCl:            50 mM
                          Tris-HCl:       10 mM
                          pH:             8.3

          Primers are available from Research Genetics Inc
          (http://www.resgen.com phone: 800-533-4363).
FEATURES            Location/Qualifiers
     source         1. .297
                    /organism="Danio rerio"
                    /strain="AB"
                    /note="Vector: ml3MP19 with added BstXI site; V-type:
                    Phage; Genomic DNA from a single adult Zebrafish of AB
                    strain was digested with AluI, Cac8I, HaeIII, NlaVI, or
                    RsaI. Fragments in the range of 250-500 bp were gel
                    purified and a BstX1 linker was added. The fragments were
                    cloned into a modified M13mp19 vector and transformed
                    into E. Coli DH5alpha. Microsatelllite sequences were
                    screened with labeled d(CA)15 and d(GT)15 oligonucleotide
                    probes."
                    /db_xref="taxon:7955"
                    /clone_lib="Zebrafish AB"
                    /sex="F"
                    /dev_stage="Adult"
                    /lab_host="DH5alphaF'IQ"
     STS            71. .200
     primer_bind    71. .96
     primer_bind    complement(180. .200)
BASE COUNT      58 a     44 c     88 g    104 t      3 others
ORIGIN

  Query Match           1.4%;  Score 68;  DB 34;  Length 297;
  Best Local Similarity  79.8%;  Pred. No. 5.24e-25;
  Matches   91;  Conservative    0;  Mismatches  23;  Indels   0;  Gaps   0;

     184 AGAAGACTACCGCTCAAAACTAGGTGGGGATTGCTTTGCTGATTTGGCCTGTCCAGAGAA 243
         |||||| || || |||||| || |||| || |||||||| ||| |||| || || || ||
Qy   1455 agaagattatcgatcaaaattaagtggagactgctttgcggatctggcttgccctgaaaa 1514

Db    244 ATGTCGATGTAAAGGAACCACAGTGGACTGCTCTGGACAAAAGCTCACCAAGAT 297
          ||||| ||| ||||||||||||| || ||||||   |||||||||| ||| ||
Qy   1515 gtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagctcaacaaaat 1568


RESULT   14
LOCUS       AF088902      721 bp    mRNA            ROD       09-MAR-1999
DEFINITION  Mus musculus SLIT1 protein (Slit1) mRNA, partial cds.
ACCESSION   AF088902
NID         g4378026
VERSION     AF088902.1  GI:4378026
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Holmes,G.P., Negus,K., Burridge,L., Raman,S., Algar,E., Yamada,T.
            and Little,M.H.
  TITLE     Distinct but overlapping expression patterns of two vertebrate slit
            homologs implies functional roles in CNS development and
            organogenesis
  JOURNAL   Mech. Dev. 79 (1-2), 57-72 (1998)
REFERENCE   2  (bases 1 to 721)
  AUTHORS   Negus,K. and Little,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-1998) Centre for Molecular and Cellular Biology,
            University of Queensland, St. Lucia, Brisbane, Qld 4072, Australia
FEATURES            Location/Qualifiers
     source         1. .721
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /db_xref="dbEST:AA423126"
     gene           <1. .721
                    /gene="Slit1"
     misc_feature   <1. .96
                    /gene="Slit1"
                    /note="Region: EGF-like repeat 8"
     CDS            <1. .336
                    /gene="Slit1"
                    /note="leucine-rich and EGF-like repeat containing
                    protein; similar to Drosophila slit"
                    /codon_start=1
                    /product="SLIT1 protein"
                    /protein_id="AAD19349.1"
                    /db_xref="PID:g4378027"
                    /db_xref="GI:4378027"
                    /translation="AAFKCHHGQCHISDRGEPYCLCQPGFSGHHCEQENPCMGEIVRE
                    AIRRQKDYASCATASKVPIMECRGGCGSQCCQPIRSKRRKYVFQCTDGSSFVEEVERH
                    LECGCRACS"
     misc_feature   196. .333
                    /gene="Slit1"
                    /note="cysteine-rich knot; unclassified site"
BASE COUNT      219 a    153 c    170 g    179 t
ORIGIN

  Query Match           1.4%;  Score 67;  DB 32;  Length 721;
  Best Local Similarity  63.0%;  Pred. No. 2.56e-24;
  Matches   199;  Conservative    0;  Mismatches  114;  Indels   3;  Gaps   3;

Db      8 TCAAGTGCCACCATGGGCAGTGTCACATATCAGATCGAGGGGAGCCCTATTGCCTATGCC 67
          |||||||| | |||||| ||||     | |||| ||  ||| |||||||| ||   ||||
Qy   4244 tcaagtgcaagcatgggaagtgcaggctttcaggtctggggcagccctactgtgaatgca 4303

Db     68 AGCCTGGCTTCAGTGGCCATCACTGTGAGCAAGAGAATCCATGTATGGGGGAGATAGTCC 127
             ||| | ||  ||  |   |||||| | ||| |   | |||   ||||| |   |
Qy   4304 gcagtggatacacgggggacagctgtgatcgagaaatctcttgtcgaggggaaaggataa 4363

Db    128 GTGAAGCCATCCGCCGCCAGAAAGACTACGCCTCTTGTGCCACGGCGTCCAAGGTGCCCA 187
          | ||      ||    ||| | | ||| ||  ||||    ||  |    |||||| ||
Qy   4364 gagattattaccaaaagcagcagggctatgctgcttgccaaacaaccaagaaggtgtccc 4423

Db    188 TCATGGAATGCCGCGG-GGGC-TGCGGGAGC-CAGTGCTGCCAGCCGATTCGAAGCAAGC 244
            | || ||| | || |||  ||| ||||  ||||||||    ||| |  | |||||||
Qy   4424 gattagagtgcagaggtgggtgtgcaggagggcagtgctgtggaccgctgaggagcaagc 4483

Db    245 GGCGGAAATATGTCTTCCAGTGCACGGACGGCTCCTCATTCGTGGAAGAGGTGGAGAGAC 304
          ||||||||||    ||| | ||||| |||||||||||| || ||||| ||||| |||| |
Qy   4484 ggcggaaatactctttcgaatgcactgacggctcctcctttgtggacgaggttgagaaag 4543

Db    305 ACTTGGAATGTGGCTG 320
              || | || |||||
Qy   4544 tggtgaagtgcggctg 4559


RESULT   15
LOCUS       I66494      7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match           1.2%;  Score 57;  DB 25;  Length 7218;
  Best Local Similarity   0.8%;  Pred. No. 1.49e-17;
  Matches    3;  Conservative  208;  Mismatches 154;  Indels   0;  Gaps   0;

Db   1064 GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1123
            ||| ::::::     :: : ::::::   ::: ::: :::::: :     :    :: :
Qy   1036 gatgctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcaca 1095

Db   1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183
             ::::::     ::: :::      :: :::::::: :  ::::: :: ::   : ::
Qy   1096 gaactccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgcc 1155

Db   1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243
             :    :   :: :::::   :   : ::::::    ::::: :  :::   :::::::
Qy   1156 aacaagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctc 1215

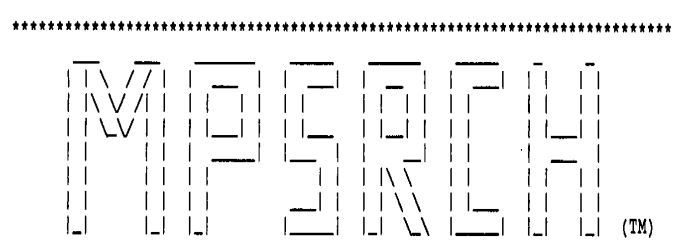
Db   1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303
            ::::: : :  :  :   ::::   :: :: ::      :::::::: ::::::::   ::
Qy   1216 tccctatatgacaacaagcttcagaccatcgccaaggggacctttttcacctcttcgggcc 1275

Db   1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
             :::   :: : : :::  :::    ::::::: ::: :  :: :: ::::   :  ::
Qy   1276 attcaaactatgcatttggcccagaaccccttttatttgtgactgccatctcaagtggcta 1335

Db   1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1423
             :   :: ::::: : ::  :::  ::    ::  :  : ::: :: : ::  ::::: :
Qy   1336 gcggattatctccataccaacccgattgagaccagtggtgcccgttgcaccagcccccgc 1395

Db   1424 YYYYY 1428
            : :::
Qy   1396 cgcct 1400


rch completed: Sat May 29 20:36:33 1999
Job time : 8593 secs.
```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

[redacted]rch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat May 29 22:50:43 1999; MasPar time 1032.12 Seconds
987.779 Million cell updates/sec
Tabular output not generated.

Title: >US-09-191-647-1
Description: (1-4758) from US09191647.seq
Perfect Score: 4758
N.A. Sequence: 1 atgcgcggcgttggctggca.....aaaaaaaaaaaaaaactcga 4758
Comp: tacgcgccgcaaccgaccgt.....ttttttttttttttgagct

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 10.449; Variance 6.610; scale 1.581

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	1231	25.9	5094	41	V16978	Nucleic acid encoding	0.00e+00
	2	200	4.2	684	41	V16966	Nucleic acid sequence	3.22e-106
	3	171	3.6	551	41	V16967	Nucleic acid sequence	2.32e-87
	4	104	2.2	8378	4	Q25811	Drosophila SLIT prote	1.07e-44
	5	54	1.1	5617	51	V57163	Partial human Notch-3	9.57e-15
	6	54	1.1	8091	50	V57001	Human Notch3 cDNA.	9.57e-15
	7	41	0.9	91	9	Q51746	Oligonucleotide probe	1.20e-07
c	8	43	0.9	91	9	Q51746	Oligonucleotide probe	1.05e-08
	9	43	0.9	204	1	N81164	Base substituted E.co	1.05e-08
c	10	44	0.9	204	1	N81164	Base substituted E.co	3.08e-09
	11	45	0.9	341	21	T24530	Human gene signature	8.92e-10
c	12	37	0.8	114	12	Q70468	Generic DNA sequence	1.41e-05
c	13	36	0.8	114	12	Q70469	Generic DNA sequence	4.52e-05
	14	32	0.7	91	46	V44650	Mammalian DNA replica	4.23e-03
c	15	31	0.7	91	46	V44650	Mammalian DNA replica	1.27e-02
	16	34	0.7	114	12	Q70467	Generic DNA sequence	4.49e-04
	17	34	0.7	114	12	Q70468	Generic DNA sequence	4.49e-04
	18	34	0.7	114	12	Q70465	Generic DNA sequence	4.49e-04
	19	34	0.7	114	12	Q70469	Generic DNA sequence	4.49e-04
	20	33	0.7	114	12	Q70466	Generic DNA sequence	1.39e-03
	21	32	0.7	114	12	Q70470	Generic DNA sequence	4.23e-03
c	22	34	0.7	114	12	Q70467	Generic DNA sequence	4.49e-04
c	23	34	0.7	114	12	Q70465	Generic DNA sequence	4.49e-04
c	24	33	0.7	114	12	Q70470	Generic DNA sequence	1.39e-03
c	25	33	0.7	114	12	Q70472	Generic DNA sequence	1.39e-03
c	26	32	0.7	114	12	Q70466	Generic DNA sequence	4.23e-03
c	27	32	0.7	168	32	T76270	Human MDNCF antisense	4.23e-03
c	28	32	0.7	172	32	T76363	Human interleukin 8 a	4.23e-03
	29	31	0.7	178	32	T76405	Human endothelin-1 an	1.27e-02
	30	31	0.7	2088	28	T58897	C-Delta-1 gene.	1.27e-02
	31	33	0.7	2446	17	T08768	Rat biglycan cDNA.	1.39e-03
	32	35	0.7	2663	35	T70174	Proliferation and dif	1.43e-04
	33	33	0.7	2692	28	T58899	M-Delta-1 gene.	1.39e-03
	34	35	0.7	2883	28	T58898	C-Delta-1 gene (alter	1.43e-04
	35	31	0.7	4208	35	T70175	Proliferation and dif	1.27e-02
	36	31	0.7	4208	40	V15201	Human serrate 1 encod	1.27e-02
	37	31	0.7	5458	39	V03674	Human Jagged gene.	1.27e-02
	38	31	0.7	6464	24	T40090	Human Serrate-1 (HJ1)	1.27e-02
	39	35	0.7	8224	2	Q12261	Versican gene.	1.43e-04
	40	30	0.6	114	12	Q70472	Generic DNA sequence	3.77e-02
	41	29	0.6	114	12	Q70471	Generic DNA sequence	1.10e-01
	42	30	0.6	1506	16	Q95302	Murine Fas antigen ex	3.77e-02
	43	30	0.6	1506	53	V71961	Fas ligand (FasL) pro	3.77e-02
	44	30	0.6	1506	21	T16305	Coding sequence for m	3.77e-02
	45	30	0.6	1515	19	T07072	Adhesive protein gene	3.77e-02

ALIGNMENTS

RESULT 1
ID V16978 standard; cDNA to mRNA; 5094 BP.
AC V16978;
DT 06-JUL-1998 (first entry)
DE Nucleic acid encoding a human slit-like polypeptide.
KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW cancer; antibody; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 233..4837
FT /*tag= a
FT sig_peptide 233..310
FT /*tag= b
FT mat_peptide 311..4834
FT /*tag= c
PN J10087699-A.
PD 07-APR-1998.
PF 15-JUL-1997; 205351.
PR 16-JUL-1996; JP-186219.
PA (ASAH) ASAHI KASEI KOGYO KK.
DR WPI; 98-267127/24.
DR P-PSDB; W46967.
PT Human Slit-like protein - useful for diagnosis and treatment of
PT brain-specific diseases and cancers
PS Disclosure; Pages 24-30; 45pp; Japanese.
CC The present sequence encodes a novel human slit-like protein (the
CC mature protein is claimed in Claim 1). The slit-like polypeptide is
CC useful for diagnosis and treatment of brain-specific diseases and
CC cancers. Antibodies directed against the protein, or its fragments
CC can also be used for diagnosing cancer.
SQ Sequence 5094 BP; 1037 A; 1616 C; 1508 G; 933 T;

Query Match 25.9%; Score 1231; DB 41; Length 5094;

Best Local Similarity 65.0%; Pred. No. 0.00e+00;
Matches 2926; Conservative 0; Mismatches 1551; Indels 24; Gaps 17;

```
Db   328 ggcgtgccccgccctctgcacctgcaccggaaccacggtggactgccacggcacggggct 387
Qy    78 ggcgtgcccggcgcagtgctcttgctcgggcagcacagtggactgtcacgggctggcgct 137

Db   388 gcaggccattcccaagaatatacctcggaacaccgagcgcctggaactcaatggcaacaa 447
Qy   138 gcgcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatggaaataa 197

Db   448 catcactcggatccatagaatgactttgcggggctcaagcagctgcgggtgctgcagct 507
Qy   198 catcacaagaattacgaagacagattttgctggtcttagacatctaagagttcttcagct 257

Db   508 gatggagaaccagattggagcagtggaacgtggtgcttttgatgacatgaaggagctgga 567
Qy   258 tatggagaataagattagcaccattgaaagaggagcattccaggatcttaaagaactaga 317

Db   568 gcggctgcgactgaaccgaaaccagctgcacatgttaccggaactgctgttccagaacaa 627
Qy   318 gagactgcgtttaaacagaaatcaccttcagctgtttcctgagttgctgtttcttgggac 377

Db   628 ccaggctttgtcaagactggacttgagtgagaacgccatccaggccatccccaggaaagc 687
Qy   378 tgcgaagctatacaggcttgatctcagtgaaaaccaaattcaggcaatcccaaggaaagc 437

Db   688 ttttcggggagctacggaccttaaaaatttacggctggacaagaaccagatcagctgcat 747
Qy   438 tttccgtggggcagttgacataaaaaatttgcaactggattacaaccagatcagctgtat 497

Db   748 tgaggaaggggccttccgtgctctgcgggggctggaggtgctgaccctgaacaacaacaa 807
Qy   498 tgaagatggggcattcagggctctccgggacctggaagtgctcactctcaacaataacaa 557

Db   808 tatcaccaccatccccgtgtccagcttcaaccatatgcccaagctacggaccttccgcct 867
Qy   558 cattactagactttctgtggcaagtttcaaccatatgcctaaacttaggacttttcgact 617

Db   868 gcactccaaccacctgttttgcgactgccacctggcctggctctcgcagtggctgaggca 927
Qy   618 gcattcaaacaacctgtattgtgactgccacctggcctggctctccgactggcttcgcaa 677

Db   928 gcggccaaccatcgggctcttcacccagtgctcgggcccagccagcctgcgtggcctcaa 987
Qy   678 aaggcctcgggttggtctgtacactcagtgtatgggcccctcccacctgagaggccataa 737

Db   988 tgtggcagaggtccagaagagtgagttcagctgctcaggccagggagaagcggggc-g-c 1045
Qy   738 tgtagccgaggttcaaaaacgagaatttgtctgcagtgatgaggaagaaggtcaccagtc 797

Db  1046 gtgcccacctgcaccctgtcctccggctc-ctgcccggccatgtgcacctgcagcaatgg 1104
Qy   798 atttatggctccttcttgtagtgttttgcactgccctgccgcctgtacctgtagcaacaa 857

Db  1105 catcgtggactgtcgtggaaaaggcctcactgccatcccggccaacctgcccgagaccat 1164
Qy   858 tatcgtagactgtcgtgggaaaggtctcactgagatccccacaaatcttccagagaccat 917

Db  1165 gacggagatacgcctggagctgaacggcatcaagtccatccctcctggagccttctcacc 1224
Qy   918 cacagaaatacgtttggaacagaacacaatcaaagtcatccctcctggagctttctcacc 977

Db  1225 ctacagaaagctacggaggatagacctgagcaacaatcagatcgctgagattgcacccga 1284
Qy   978 atataaaaagcttagacgaattgacctgagcaataatcagatctctgaacttgcaccaga 1037

Db  1285 cgccttccagggcctccgctccctgaactcgctggtcctctatggaaacaagatcacaga 1344
Qy  1038 tgctttccaaggactacgctctctgaattcacttgtcctctatggaaataaaatcacaga 1097

Db  1345 cctcccccgtggtgtgtttggaggcctatacaccctacagctcctgctcctgaatgccaa 1404
Qy  1098 actccccaaaagtttatttgaaggactgttttccttacagctcctattattgaatgccaa 1157

Db  1405 caagatcaactgcatccggcccgatgccttccaggacctgcagaacctctcactgctctc 1464
Qy  1158 caagataaactgccttcgggtagatgcttttcaggatctccacaacttgaaccttctctc 1217

Db  1465 cctgtatgacaacaagatccagagcctcgccaagggcactttcacctccctgcgggccat 1524
Qy  1218 cctatatgacaacaagcttcagaccatcgccaaggggacctttcacctcttcgggccat 1277

Db  1525 ccagactctgcacctggcccagaacccttttatttgtgactgccatctcaagtggctggc 1584
Qy  1278 tcaaactatgcatttggcccagaacccctttatttgtgactgccatctcaagtggctagc 1337

Db  1585 agacttcctgcgcaccaatcccatcgagacgagtggtgcccgctgtgccagtccccggcg 1644
Qy  1338 ggattatctccatacccaacccgattgagaccagtggtgcccgttgcaccagcccccgccg 1397

Db  1645 cctcgccaacaagcgcatcgggcagatcaagagcaagaagttccggtgctcagccaaaga 1704
Qy  1398 cctggcaaacaaaagaattggacagatcaaaagcaagaaattccgttgttcaggtacaga 1457

Db  1705 gcagtacttcattccaggcacggaggattaccagctgaacagcgagtgcaacagcgacgt 1764
Qy  1458 --ag-a-tt-at-cgat-caa--a--attaagtggaga-ctgctt-tgc----g-gatct 1499

Db  1765 ggtctgtccccacaagtgccgctgtgaggccaacgtggtggagtgctccagcctgaagct 1824
Qy  1500 ggcttgccctgaaaagtgtcgctgtgaaggaaccacagtagattgctctaatcaaaagct 1559

Db  1825 caccaagatccctgagcgcatcccccagtccacggcagaactgcgattgaataacaatga 1884
Qy  1560 caacaaaatcccggagcacattccccagtacactgcagagttgcgtctcaataataatga 1619

Db  1885 gatttccatcctggaggccactgggatgtttaaaaaacttacacatctgaagaaaatcaa 1944
Qy  1620 atttaccgtgttggaagccacaggaatctttaagaaacttcctcaattacgtaaaataaa 1679

Db  1945 tctgagcaacaacaaggtgtcagaaattgaagatggggccttcgagggcgcagcctctgt 2004
Qy  1680 ctttagcaacaataagatcacagatattgaggagggagcatttgaaggagcatctggtgt 1739

Db  2005 gagcgagctgcacctaactgccaaccagctggagtccatccggagcggcatgttccgggg 2064
Qy  1740 aaatgaaatacttcttacgagtaatcgtttggaaaatgtgcagcataagatgttcaaggg 1799

Db  2065 tctggatggcttgaggaccctaatgctgcggaacaaccgcatcagctgcatccacaacga 2124
Qy  1800 attggaaagcctcaaaactttgatgttgagaagcaatcgaataacctgtgtggggaatga 1859

Db  2125 cagcttcacgggcctgcgcaacgtccggctcctctcgctctacgacaaccagatcaccac 2184
Qy  1860 cagtttcataggactcagttctgtgcgtttgctttctttgtatgataatcaaattactac 1919

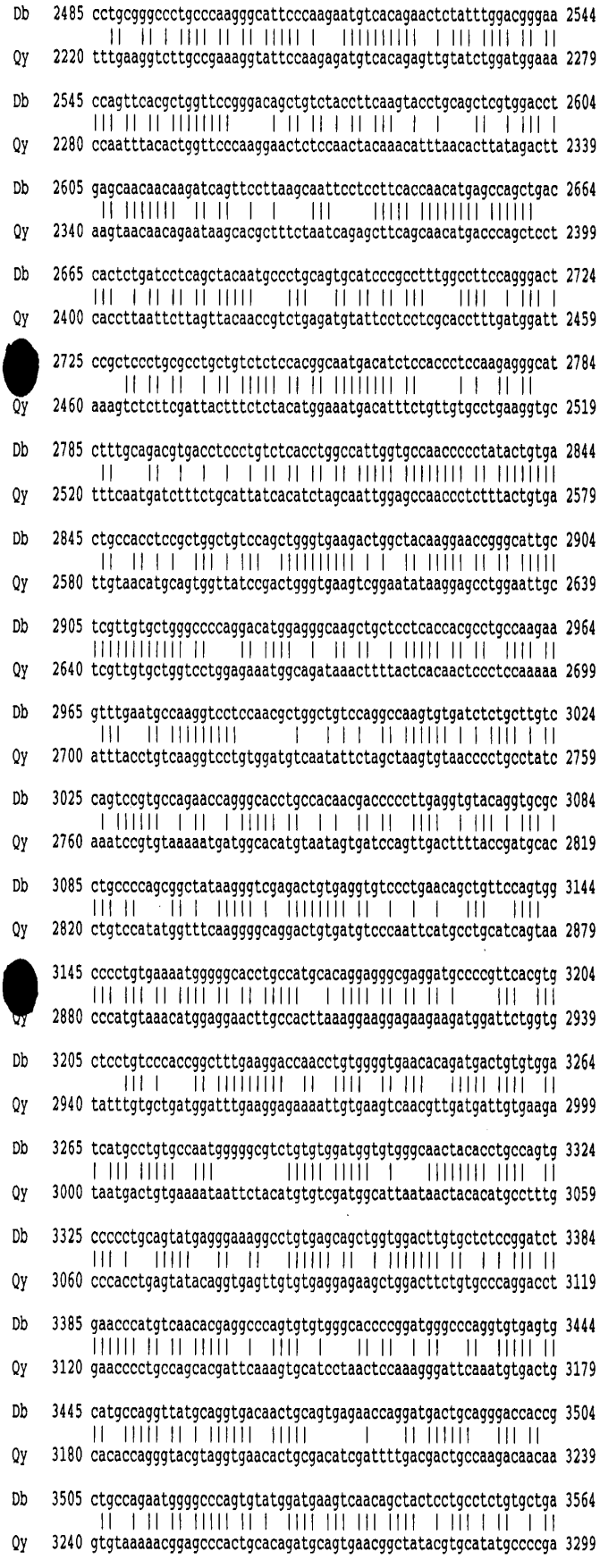
Db  2185 cgtatccccaggagccttcgacaccctccagtccctctccacactgaatctcctggccaa 2244
Qy  1920 agttgcaccaggggcatttgatactctccattctttatctactctaaacctcttggccaa 1979

Db  2245 ccctttcaactgcaactgccagctggcctggctaggaggctggctacggaagcgcaagat 2304
Qy  1980 tccttttaactgtaactgctacctggcttggttgggagagtggctgagaaagaagagaat 2039

Db  2305 cgtgacggggaacccgcgatgccagaaccctgacttttgcggcagattcccctgcagga 2364
Qy  2040 tgtcacgggaaatcctagatgtcaaaaaccatacttcctgaaagaaatacccatccagga 2099

Db  2365 cgtggccttccctgacttcaggtgtgaggaaggccaggaggaggggggctgcctgccccg 2424
Qy  2100 tgtggccattcaggacttcacttgtgatgacggaaatgatgacaatagttgctccccact 2159

Db  2425 cccacagtgcccacaggagtgcgcctgcctggacaccgtggtccgatgcagcaacaagca 2484
Qy  2160 ttctcgctgtcctactgaatgtacttgcttggatacagtcgtccgatgtagcaacaaggg 2219
```


Db 2485 cctgcgggccctgcccaagggcattcccaagaatgtcacagaactctatttggacgggaa 2544
Qy 2220 tttgaaggtcttgccgaaaggtattccaagagatgtcacagagttgtatctggatggaaa 2279
Db 2545 ccagttcacgctggttccgggacagctgtctaccttcaagtacctgcagctcgtggacct 2604
Qy 2280 ccaatttacactggttcccaaggaactctccaactacaaacatttaacacttatagactt 2339
Db 2605 gagcaacaacaagatcagttccttaagcaattcctccttcaccaacatgagccagctgac 2664
Qy 2340 aagtaacaacagaataagcacgctttctaatcagagcttcagcaacatgacccagctcct 2399
Db 2665 cactctgatcctcagctacaatgccctgcagtgcatcccgcctttggccttccagggact 2724
Qy 2400 caccttaattcttagttacaaccgtctgagatgtattcctcctcgcacctttgatggatt 2459
2725 ccgctccctgcgcctgctgtctctccacggcaatgacatctccaccctccaagagggcat 2784
Qy 2460 aaagtctcttcgattactttctctacatggaaatgacatttctgttgtgcctgaaggtgc 2519
Db 2785 ctttgcagacgtgacctccctgtctcacctggccattggtgccaaccccctatactgtga 2844
Qy 2520 tttcaatgatctttctgcattatcacatctagcaattggagccaaccctctttactgtga 2579
Db 2845 ctgccacctccgctggctgtccagctgggtgaagactggctacaaggaaccgggcattgc 2904
Qy 2580 ttgtaacatgcagtggttatccgactgggtgaagtcggaatataaggagcctggaattgc 2639
Db 2905 tcgttgtgctgggccccaggacatggagggcaagctgctcctcaccacgcctgccaagaa 2964
Qy 2640 tcgttgtgctggtcctggagaaatggcagataaacttttactcacaactccctccaaaaa 2699
Db 2965 gtttgaatgccaaggtcctccaacgctggctgtccaggccaagtgtgatctctgcttgtc 3024
Qy 2700 atttacctgtcaaggtcctgtggatgtcaatattctagctaagtgtaacccctgcctatc 2759
Db 3025 cagtccgtgccagaaccagggcacctgccacaacgacccccttgaggtgtacaggtgcgc 3084
Qy 2760 aaatccgtgtaaaaatgatggcacatgtaatagtgatccagttgacttttaccgatgcac 2819
Db 3085 ctgccccagcggctataagggtcgagactgtgaggtgtccctgaacagctgttccagtgg 3144
Qy 2820 ctgtccatatggtttcaaggggcaggactgtgatgtcccaattcatgcctgcatcagtaa 2879
3145 cccctgtgaaaatgggggcacctgccatgcacaggagggcgaggatgccccgttcacgtg 3204
Qy 2880 cccatgtaaacatggaggaacttgccacttaaaggaaggagaagaagatggattctggtg 2939
Db 3205 ctcctgtcccaccggctttgaaggaccaacctgtggggtgaacacagatgactgtgtgga 3264
Qy 2940 tatttgtgctgatggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaaga 2999
Db 3265 tcatgcctgtgccaatgggggcgtctgtgtggatggtgtgggcaactacacctgccagtg 3324
Qy 3000 taatgactgtgaaaataattctacatgtgtcgatggcattaataactacacatgcctttg 3059
Db 3325 ccccctgcagtatgagggaaaggcctgtgagcagctggtggacttgtgctctccggatct 3384
Qy 3060 cccacctgagtatacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacct 3119
Db 3385 gaacccatgtcaacacgaggcccagtgtgtgggcaccccggatgggcccaggtgtgagtg 3444
Qy 3120 gaacccctgccagcacgattcaaagtgcatcctaactccaaagggattcaaatgtgactg 3179
Db 3445 catgccaggttatgcaggtgacaactgcagtgagaaccaggatgactgcagggaccaccg 3504
Qy 3180 cacaccagggtacgtaggtgaacactgcgacatcgattttgacgactgccaagacaacaa 3239
Db 3505 ctgccagaatggggcccagtgtatggatgaagtcaacagctactcctgcctctgtgctga 3564
Qy 3240 gtgtaaaaacggagcccactgcacagatgcagtgaacggctatacgtgcatatgccccga 3299
Db 3565 gggctacagtggacagctctgtgagatccctcc--ccatctgcctgcccccaag-agccc 3621
Qy 3300 aggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtaccagccc 3359
Db 3622 ctgtgaggggactgagtgccagaatggggccaactgtgtggaccagggcaacaggcctgt 3681
Qy 3360 ctgtgataatattttgattgtcagaatggagctcagtgtatcgtcagaataaatgagccaat 3419
Db 3682 gtgccagtgcctcccaggcttcggtggccctgagtgtgagaagttgctcagtgtcaactt 3741
Qy 3420 atgtcagtgtttgcctggctatcagggagaaaagtgtgaaaaattggttagtgtgaatttt 3479
Db 3742 tgtggatcgggacacttacctgcagttcactgacctgcaaaactggccacgggccaacat 3801
Qy 3480 tataaacaaagagtcttatcttcagattccttcagccaaggttcggcctcagacgaacat 3539
Db 3802 cacgttgcaggtctccacggcagaggacaatgggatccttctgtacaacggggacaacga 3861
Qy 3540 aacacttcagattgccacagatgaagacagcggaatcctcctgtataagggtgacaaaga 3599
Db 3862 ccacattgcagttgagctgtaccagggccatgtgcgtgtcagctacgacccaggcagcta 3921
Qy 3600 ccatatcgcggtagaactctatcgggggcgtgttcgtgccagctatgacaccggctctca 3659
Db 3922 ccccagctctgccatctacagtgctgagacgatcaacgatgggcaattccacaccgttga 3981
Qy 3660 tccagcttctgccatttacagtgtggagacaatcaatgatggaaacttccacattgtgga 3719
Db 3982 gctggttgcctttgaccagatggtgaatctctccattgatggcgggagccccatgaccat 4041
Qy 3720 actacttgccttggatcagagtctctctttgtccgtggatggtgggaaccccaaaatcat 3779
Db 4042 ggacaactttggcaaacattacacgctcaacagcgaggcgccactctatgtgggagggat 4101
Qy 3780 cactaacttgtcaaagcagtccactctgaattttgactctccactctatgtaggaggcat 3839
Db 4102 gcccgtggatgtcaactcagctgccttccgcctgtggcagatcctcaacggcaccggctt 4161
Qy 3840 gccagggaagagtaacgtggcatctctgcgccaggcccctgggcagaacggaaccagctt 3899
Db 4162 ccacggttgcatccgaaacctgtacatcaacaacgagctgcaggacttcaccaagacgca 4221
Qy 3900 ccacggctgcatccggaaccttacatcaacagtgagctgcaggacttccagaaggtgcc 3959
Db 4222 gatgaagccaggcgtggtgccaggctgcgaaccctgccgcaagctctactgcctgcatgg 4281
Qy 3960 gatgcaaacaggcattttgcctggctgtgagccatgccacaagaaggtgtgtgcccatgg 4019
Db 4282 catctgccagcccaatgccaccccagggcccatgtgccactgcgaggctggctgggtggg 4341
Qy 4020 cacatgccagcccagcagccaggcaggcttcacctgcgagtgccaggaaggatggatggg 4079
Db 4342 cctgcactgtgaccagcccgctgacggcccctgccatggccacaagtgtgtccatgggca 4401
Qy 4080 gcccctctgtgaccaacggaccaatgacccttgccttggaaataaatgcgtacatggcac 4139
Db 4402 atgcgtgcccctcgacgctctttcctacagctgccagtgccaggatgggtactcgggggc 4461
Qy 4140 ctgcttgcccatcaatgcgttctcctacagctgtaagtgcttggagggccatggaggtgt 4199
Db 4462 actgtgcaaccaggccggggccctggcagagccctgcagaggcctgcagtgcctgcatgg 4521
Qy 4200 cctctgtgatgaagaggaggatctgtttaacccatgccaggcgatcaagtgcaagcatgg 4259
Db 4522 ccactgccaggcctcaggcaccaaggggcacactgtgtgtgtgaccccggcttttcggg 4581
Qy 4260 gaagtgcaggctttcaggtctggggcagccctactgtgaatgcagcagtggatacacggg 4319
Db 4582 cgagctgtgtgagcaagagtccgagtgccgggggggaccctgtccgggactttcaccaggt 4641
Qy 4320 ggacagctgtgatcgagaaatctcttgtcgaggggaaaggataagagattattaccaaaa 4379
Db 4642 ccagaggggctatgccatctgccagaccacgcgccccctgtcatgggtggagtgccgggg 4701

```
Qy   4380 gcagcagggctatgctgcttgccaaacaaccaagaaggtgtcccgattagagtgcagagg 4439

Db   4702 ctcgtgcccaggccagggctgctgccagggccttcggctgaagcggaggaagttcaccttt 4761
Qy   4440 tgggtgtgcaggagggcagtgctgtggaccgctgaggagcaagcggcggaaatactcttt 4499

Db   4762 tgagtgcagcgatgggacctcttttgccgaggaggtggaaaagcccaccaagtgtggctg 4821
Qy   4500 cgaatgcactgacggctcctcctttgtggacgaggttgagaaagtggtgaagtgcggctg 4559

Db   4822 t 4822
Qy   4560 t 4560


RESULT    2
ID   V16966 standard; cDNA to mRNA; 684 BP.
AC   V16966;
DT   06-JUL-1998  (first entry)
DE   Nucleic acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody; ds.
OS   Mus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   P-PSDB; W46967.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Pages 35-36; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   684 BP;    143 A;    206 C;    199 G;    136 T;

  Query Match           4.2%;  Score 200;  DB 41;  Length 684;
  Best Local Similarity  66.0%;  Pred. No. 3.22e-106;
  Matches   413;  Conservative    0;  Mismatches 213;  Indels   0;  Gaps   0;

Db      4 gaggacaatgggattctcctatataatggggacaatgaccacattgcagttgagctgtac 63
Qy   3562 gaagacagcggaatcctcctgtataagggtgacaaagaccatatcgcggtagaactctat 3621

Db     64 cagggccatgtccgtgttagttatgacccaggcagctaccccagctctgctatctacagc 123
Qy   3622 cgggggcgtgttcgtgccagctatgacaccggctctcatccagcttctgccatttacagt 3681

Db    124 gctgagacgatcaacgatgggcagttccacaccgttgagctggtcaccttcgaccagatg 183
Qy   3682 gtggagacaatcaatgatggaaacttccacattgtggaactacttgccttggatcagagt 3741

Db    184 gtgaacctctccatcgatggcggcagccccatgaccatggacaactttggcaagcactac 243
Qy   3742 ctctctttgtccgtggatggtgggaaccccaaaatcatcactaacttgtcaaagcagtcc 3801

Db    244 accctcaacagcgaggccccccctctatgtgggagggatgcctgtggacgtgaactcagct 303
Qy   3802 actctgaattttgactctccactctatgtaggaggcatgccagggaagagtaacgtggca 3861

Db    304 gccttccgcctgtggcagatcctcaatggtaccagcttccacggttgcatccggaacctg 363
Qy   3862 tctctgcgccaggcccctgggcagaacggaaccagcttccacggctgcatccggaacctt 3921

Db    364 tacatcaacaatgaactgcaggacttcaccaagacacagatgaagccaggcgtggtgcca 423
Qy   3922 tacatcaacagtgagctgcaggacttccagaaggtgccgatgcaaacaggcattttgcct 3981

Db    424 ggctgtgagccctgccgcaaactctactgtctacatggcatttgccagcccaacgccacc 483
Qy   3982 ggctgtgagccatgccacaagaaggtgtgtgcccatggcacatgccagcccagcagccag 4041

Db    484 ccagggcctgtgtgccactgtgaggctggctgggggggcctgcactgtgaccagccagtg 543
Qy   4042 gcaggcttcacctgcgagtgccaggaaggatggatggggcccctctgtgaccaacggacc 4101

Db    544 gatggcccctgccatggccacaagtgtgtccatgggaaatgcgtgccgctcgacgctctt 603
Qy   4102 aatgacccttgccttggaaataaatgcgtacatggcacctgcttgcccatcaatgcgttc 4161

Db    604 gcctacagctgccagtgccaggatgg 629
Qy   4162 tcctacagctgtaagtgcttggaggg 4187


RESULT    3
ID   V16967 standard; cDNA to mRNA; 551 BP.
AC   V16967;
DT   06-JUL-1998  (first entry)
DE   Nucleic acid sequence of the specification.
KW   Slit-like protein; human; diagnosis; treatment; brain-specific disease;
KW   cancer; antibody; ds.
OS   Rattus sp.
PN   J10087699-A.
PD   07-APR-1998.
PF   15-JUL-1997; 205351.
PR   16-JUL-1996; JP-186219.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 98-267127/24.
DR   P-PSDB; W46968.
PT   Human Slit-like protein - useful for diagnosis and treatment of
PT   brain-specific diseases and cancers
PS   Disclosure; Page 37; 45pp; Japanese.
CC   The present sequence appears in the specification. The specification
CC   describes a novel human slit-like protein (the mature protein is claimed
CC   in Claim 1). The slit-like polypeptide is useful for diagnosis and
CC   treatment of brain-specific diseases and cancers. Antibodies directed
CC   against the protein, or its fragments can also be used for diagnosing
CC   cancer.
SQ   Sequence   551 BP;    121 A;    175 C;    149 G;    106 T;

  Query Match           3.6%;  Score 171;  DB 41;  Length 551;
  Best Local Similarity  65.6%;  Pred. No. 2.32e-87;.
  Matches   360;  Conservative    0;  Mismatches 189;  Indels   0;  Gaps   0;

Db      2 gggccatgtccgtgttagctacgacccaggcagctaccccagctctgctatctacagtgc 61
Qy   3624 ggggcgtgttcgtgccagctatgacaccggctctcatccagcttctgccatttacagtgt 3683

Db     62 tgaaacaatcaacgatgggcagttccacacagttgagctggtgacctttgaccagatggt 121
Qy   3684 ggagacaatcaatgatggaaacttccacattgtggaactacttgccttggatcagagtct 3743

Db    122 gaacctctccatcgatggtggcagccccatgaccatggacaactttggaaagcactacac 181
Qy   3744 ctctttgtccgtggatggtgggaaccccaaaatcatcactaacttgtcaaagcagtccac 3803

Db    182 actcaacagtgaggccccccctctatgtgggagggatgcccgtggatgtgaactcagctgc 241
Qy   3804 tctgaattttgactctccactctatgtaggaggcatgccagggaagagtaacgtggcatc 3863

Db    242 cttccgcctgtggcagatcctcaatggcaccagcttccacggttgcatccggaatctata 301
Qy   3864 tctgcgccaggcccctgggcagaacggaaccagcttccacggctgcatccggaaccttta 3923

Db    302 catcaacaacgaactgcaggacttcaccaagacacagatgaagccgggcgtggtgcccgg 361
Qy   3924 catcaacagtgagctgcaggacttccagaaggtgccgatgcaaacaggcattttgcctgg 3983

Db    362 ctgcgagccctgccgaaaactctactgtctacatggcatttgccagcccaacgccacccc 421
```

```
Qy   3984 ctgtgagccatgccacaagaaggtgtgtgcccatggcacatgccagcccagcagccaggc 4043

Db    422 agggcccgtgtgccactgcgaggctggctgggggggcctgcactgtgaccagccagtgga 481
Qy   4044 aggcttcacctgcgagtgccaggaaggatggatggggccccctctgtgaccaacggaccaa 4103

Db    482 cggcccctgccatggccacaagtgtgtccatgggaaatgcgtgcccctcgatgcacttgc 541
Qy   4104 tgacccttgccttggaaataaatgcgtacatggcacctgcttgcccatcaatgcgttctc 4163

Db    542 ctacagctg 550
Qy   4164 ctacagctg 4172


RESULT    4
     Q25811 standard; cDNA; 8378 BP.
     Q25811;
DT   05-JAN-1993  (first entry)
DE   Drosophila SLIT protein involved in axon pathway development.
KW   Neurogenesis; EGF-like repeats; epidermal growth factor; TAGON;
KW   embryonic CNS; leucine-rich repeat; Flank-LRR-Flank; glial cells;
KW   Notch; axonogenesis; cell-cell interaction; ss.
OS   Drosophila melanogaster.
FH   Key             Location/Qualifiers
FT   5'utr           1..314
FT                   /*tag= a
FT   cds             315..4757
FT                   /*tag= b
FT                   /product= SLIT_protein
FT   3'utr           4755..8378
FT                   /*tag= c
PN   WO9210518-A.
PD   25-JUN-1992.
PF   27-NOV-1991; U09055.
PR   07-DEC-1990; US-624135.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-Tsakonas S, Rothberg JM;
DR   WPI; 92-234590/28.
DR   P-PSDB; R25079.
PT   SLIT protein and sequence elements for treating
PT   neuro-degenerative disease - useful for Alzheimer's disease,
PT   nerve damage and Parkinson's disease, for diagnosis of cancer
PS   Claim 2; Page 74-83; 122pp; English.
CC   PCR and standard library screening techniques were used to further
CC   analyse previously isolated slit cDNA clones. A cDNA clone
CC   representing the 5' most 2.4kb of sequence was isolated from a
CC   larval library and PCR was used to isolate a corresp. sequence from
CC   a 4-8 hour embryonic library. Two forms of the SLIT message were
CC   identified differing by 33 nucleotides. Genomic and cDNA sequencing
CC   indicates the transcripts consists of an approximately 314bp 5'
CC   untranslated leader sequence, followed by either a 4407bp or 4440bp
CC   ORF depending on the splice form and a 4kb 3'-UTR.
SQ   Sequence   8378 BP;    2192 A;    2272 C;    2164 G;    1742 T;

  Query Match           2.2%;  Score 104;  DB 4;  Length 8378;
  Best Local Similarity  59.4%;  Pred. No. 1.07e-44;
  Matches   330;  Conservative    0;  Mismatches 226;  Indels   0;  Gaps   0;

Db   1196 ctgtccgcacccatgtcgctgtgcggacgggatcgtcgattgccgtgagaagagtctgac 1255
Qy    828 ctgccctgccgcctgtacctgtagcaacaatatcgtagactgtcgtgggaaaggtctcac 887

Db   1256 cagcgtgcccgtcaccttgcccgacgacaccaccgacgttcgcctcgagcaaaatttcat 1315
Qy    888 tgagatccccacaaatcttccagagaccatcacagaaatacgtttggaacagaacacaat 947

Db   1316 tacggaactgccgccgaaatcgttctccagctttcgacgactgcgacgcatcgacctgtc 1375
Qy    948 caaagtcatccctcctggagctttctcaccatataaaaagcttagacgaattgacctgag 1007

Db   1376 caacaacaacatatcccggattgcccacgatgcactaagcggcctaaagcagttaaccac 1435

Qy   1008 caataatcagatctctgaacttgcaccagatgctttccaaggactacgctctctgaattc 1067

Db   1436 tctcgtgctgtacggcaataaaataaaggatttaccctcgggcgtgttcaaaggactcgg 1495
Qy   1068 acttgtcctctatggaaataaaatcacagaactccccaaaagtttatttgaaggactgtt 1127

Db   1496 ctcgctcaggctgctgctgctgaacgccaacgagatctcgtgcatacgcaaggatgcctt 1555
Qy   1128 ttccttacagctcctattattgaatgccaacaagataaactgccttcgggtagatgcttt 1187

Db   1556 tcgcgacctgcacagtttgagcctgctctccctgtacgacaacaacatccagtcggtggc 1615
Qy   1188 tcaggatctccacaacttgaaccttctctccctatatgacaacaagcttcagaccatcgc 1247

Db   1616 taatggcacattcgacgccatgaagagcatgaaaacggtacatctggccaagaatccttt 1675
Qy   1248 caaggggacctttcacctcttcgggccattcaaactatgcatttggcccagaaccccctt 1307

Db   1676 catctgcgactgcaatctgcgctgggtggccgactatttgcacaaaaatcccatagagac 1735
Qy   1308 tatttgtgactgccatctcaagtggctagcggattatctccataccaacccgattgagac 1367

Db   1736 gagtggagcccgctgc 1751
Qy   1368 cagtggtgcccgttgc 1383


RESULT    5
ID   V57163 standard; DNA; 5617 BP.
AC   V57163;
DT   06-JAN-1999  (first entry)
DE   Partial human Notch-3 gene.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1..5616
FT                   /*tag= a
FT                   /product= "partial human Notch-3"
FT                   /transl_except= (pos: 982..984, aa: Xaa)
FT                   /transl_except= (pos: 1201..1203, aa: Xaa)
FT                   /transl_except= (pos: 1207..1209, aa: Xaa)
FT                   /transl_except= (pos: 1216..1218, aa: Xaa)
FT                   /transl_except= (pos: 1225..1227, aa: Xaa)
FT                   /transl_except= (pos: 1258..1260, aa: Xaa)
FT                   /transl_except= (pos: 2116..2118, aa: Xaa)
FT                   /transl_except= (pos: 2182..2184, aa: Xaa)
FT                   /transl_except= (pos: 2275..2367, aa: Xaa)
FT                   /transl_except= (pos; 4273..4275, aa: Xaa)
FT                   /note= "Xaa = unknown; no start or stop codons are given
FT                         at the 5' or 3' ends of the sequence"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996; 009733.
PR   01-AUG-1996; FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR   WPI; 98-133137/13.
DR   P-PSDB; W68510.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents the sequence encoding a partial human notch3
CC   protein.  Notch3 is a transmembrane receptor protein involved in lateral
CC   inhibition and regulating developmental cascades of neurogenic genes.
CC   Mutated Notch3 proteins are thought to be involved in neurological
CC   disorders, especially of the cerebral autosomal dominant arteriopathy
CC   with subcortical infarcts and leukoencephalopathy (CADASIL) type.
CC   Blocking expression of a mutated Notch3 gene or by substitution therapy
```

```
CC  with non-mutated Notch3 gene or protein can be used to treat CADASIL or
CC  related disorders.
SQ  Sequence  5617 BP;  859 A;  1824 C;  1790 G;  1033 T;

  Query Match           1.1%;  Score 54;  DB 51;  Length 5617;
  Best Local Similarity  65.0%;  Pred. No. 9.57e-15;
  Matches  117;  Conservative   0;  Mismatches  63;  Indels   0;  Gaps   0;

Db   480 tgggtttgagggtcagaattgtgaagtgaacgtggacgactgtccaggacaccgatgtct 539
         ||| ||||| ||  | ||||||||||| ||||| || || |||  ||   |    |||
Qy  2952 tggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaagataatgactgtga 3011

Db   540 caatgggggggacatgcgtggatggcgtcaacacctataactgccagtgccctcctgagtg 599
         |||      ||||| || |||||| | || | ||| |  ||||  ||||| |||||||
Qy  3012 aaataattctacatgtgtcgatggcattaataactacacatgcctttgcccacctgagta 3071

Db   600 gacaggccagttctgcacggaggacgtggatgagtgtcagctgcagcccaacgcctgcca 659
         |||||  |||| ||   |||| |  ||||    |||   | | | |  ||| |||||||
Qy  3072 tacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacctgaacccctgcca 3131


RESULT  6
V57001 standard; cDNA; 8091 BP.
AC  V57001;
DT  21-DEC-1998  (first entry)
DE  Human Notch3 cDNA.
KW  Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW  developmental cascade; neurogenic gene; mutant; neurological disorder;
KW  cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW  leukoencephalopathy; therapy; ss.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  CDS             79..7044
FT                  /*tag= a
FT                  /product= "Notch3 protein"
PN  FR2751986-A1.
PD  06-FEB-1998.
PF  16-APR-1997; 004680.
PR  01-AUG-1996; FR-009733.
PA  (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI  Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
DR  WPI; 98-133138/13.
DR  P-PSDB; W49698.
PT  Human Notch3 nucleic acids - and methods for identifying
PT  pre-disposition to cerebral autosomal dominant arteriopathy with
PT  sub-cortical infarcts and leukoencephalopathy
PS  Claim 2; Fig 1.1-1.8; 45pp; French.
CC  This sequence represents the cDNA sequence encoding the human notch3
CC  protein. Notch3 is a transmembrane receptor protein involved in lateral
CC  inhibition and regulating developmental cascades of neurogenic genes.
CC  Mutated Notch3 proteins are thought to be involved in neurological
CC  disorders, especially of the cerebral autosomal dominant arteriopathy
CC  with subcortical infarcts and leukoencephalopathy (CADASIL) type.
CC  Blocking expression of a mutated Notch3 gene or by substitution therapy
CC  with non-mutated Notch3 gene or protein can be used to treat CADASIL or
CC  related disorders.
SQ  Sequence  8091 BP;  1311 A;  2716 C;  2488 G;  1576 T;

  Query Match           1.1%;  Score 54;  DB 50;  Length 8091;
  Best Local Similarity  65.0%;  Pred. No. 9.57e-15;
  Matches  117;  Conservative   0;  Mismatches  63;  Indels   0;  Gaps   0;

Db   756 tgggtttgagggtcagaattgtgaagtgaacgtggacgactgtccaggacaccgatgtct 815
         ||| ||||| ||  | ||||||||||| ||||| || || |||  ||   |    |||
Qy  2952 tggatttgaaggagaaaattgtgaagtcaacgttgatgattgtgaagataatgactgtga 3011

Db   816 caatgggggggacatgcgtggatggcgtcaacacctataactgccagtgccctcctgagtg 875
         |||      ||||| || |||||| | || | ||| |  ||||  ||||| |||||||
Qy  3012 aaataattctacatgtgtcgatggcattaataactacacatgcctttgcccacctgagta 3071

Db   876 gacaggccagttctgcacggaggacgtggatgagtgtcagctgcagcccaacgcctgcca 935
         |||||  |||| ||   |||| |  ||||    |||   | | | |  ||| |||||||
Qy  3072 tacaggtgagttgtgtgaggagaagctggacttctgtgcccaggacctgaacccctgcca 3131


RESULT  7
ID  Q51746 standard; cDNA; 91 BP.
AC  Q51746;
DT  31-MAY-1994  (first entry)
DE  Oligonucleotide probe MK14-A
KW  Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW  ss.
OS  Synthetic.
PN  EP-571911-A.
PD  01-DEC-1993.
PF  24-MAY-1993; 108325.
PR  26-MAY-1992; US-889651.
PA  (BECT ) BECTON DICKINSON CO.
PI  Shank DD,  Spears PA;
DR  WPI; 93-378844/48.
PT  New oligo:nucleotide probes specific for Mycobacteria - used for
PT  detection and amplification of Mycobacteria nucleic acid in
PT  samples
PS  Claim 3; Page 14; 23pp; English.
CC  Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC  (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC  cross reacted to a few non-mycobacterial spp.  The probe may
CC  be useful as an initial screen for mycobacterial infection.
CC  See also Q51735-45 and Q51747-59.
SQ  Sequence  91 BP;  5 A;  17 C;  15 G;  4 T;

  Query Match           0.9%;  Score 41;  DB 9;  Length 91;
  Best Local Similarity  3.8%;  Pred. No. 1.20e-07;
  Matches    2;  Conservative   45;  Mismatches   6;  Indels   0;  Gaps   0;

Db     8 gcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhyvysv 60
         || ::::::::: :: :::::::::::::::::::: ::::::: : ::::::
Qy   140 gcagcgtgcccaggaatatcccccgcaacaccgagagactggatttaaatgga 192


RESULT  8
ID  Q51746 standard; cDNA; 91 BP.
AC  Q51746;
DT  31-MAY-1994  (first entry)
DE  Oligonucleotide probe MK14-A
KW  Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW  ss.
OS  Synthetic.
PN  EP-571911-A.
PD  01-DEC-1993.
PF  24-MAY-1993; 108325.
PR  26-MAY-1992; US-889651.
PA  (BECT ) BECTON DICKINSON CO.
PI  Shank DD,  Spears PA;
DR  WPI; 93-378844/48.
PT  New oligo:nucleotide probes specific for Mycobacteria - used for
PT  detection and amplification of Mycobacteria nucleic acid in
PT  samples
PS  Claim 3; Page 14; 23pp; English.
CC  Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC  (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC  cross reacted to a few non-mycobacterial spp.  The probe may
CC  be useful as an initial screen for mycobacterial infection.
CC  See also Q51735-45 and Q51747-59.
SQ  Sequence  91 BP;  5 A;  17 C;  15 G;  4 T;

  Query Match           0.9%;  Score 43;  DB 9;  Length 91;
  Best Local Similarity  21.3%;  Pred. No. 1.05e-08;
  Matches   16;  Conservative   43;  Mismatches  16;  Indels   0;  Gaps   0;

Db     3 ctccggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhyvysvct 62
         || ||| |::::::: :::: :::::::::::::: ::::: ::::  ::::::: :
Cp    78 ctgcggtgccaccttgttcaggatcgccagcactaaccccagcgacagggacagcatctg 19

Db    63 caagcctcggcggcg 77
         | ||||  || |||
```

```
Cp    18 ccagccaacgccgcg 4


RESULT   9
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key            Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag= a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag= b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           0.9%;  Score 43;  DB 1;  Length 204;
  Best Local Similarity  8.2%;  Pred. No. 1.05e-08;
  Matches    8;  Conservative   55;  Mismatches  34;  Indels   0;  Gaps   0;

Db     92 hhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnnc 151
           :: :: :: ::::: :::   ::  :::: :::| : ||:::   :::::: :  :
Qy   3296 ccgaaggttacagtggcttgttctgtgagttttctccacccatggtcctccctcgtacca 3355

Db    152 ncccbnnhvchnvhbnnhrnwayvrhdarrddvhccv 188
            |||:  :: : :::   : : ::: :| ::::: |:
Qy   3356 gcccctgtgataattttgattgtcagaatggagctca 3392


RESULT   10
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key            Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag= a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag= b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           0.9%;  Score 44;  DB 1;  Length 204;
  Best Local Similarity  17.6%;  Pred. No. 3.08e-09;
  Matches   30;  Conservative   72;  Mismatches  67;  Indels   1;  Gaps   1;

Db     16 acrccaagcyyrcayrccbgcaggycgacbcyrraggnyccccgggywccgagcycgaay 75
          ||:|  | |::: |:: |:||| :|  | |:       || ||  ::|  ||   || :
Cp   4573 acacacacctcgtacagccgcacttcaccactttctcaacctcgtccacaaaggaggagc 4514

Db     76 ycdchvgccgymrtttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddh 135
          : :|::  | :     :: :: ::  : :: :::   ::| ::::::: | :  |:::::
Cp   4513 cgtcagtgcattcga-aagagtatttccgccgcttgctcctcagcggtccacagcactgc 4455

Db    136 yvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvh 185
          :::::::: :: :  |  | :  : |: :::   : : :  :: ::::::
Cp   4454 cctcctgcacacccacctctgcactctaatcgggacaccttcttggttgt 4405


RESULT   11
ID   T24530 standard; cDNA to mRNA; 341 BP.
AC   T24530;
DT   25-SEP-1996  (first entry)
DE   Human gene signature HUMGS06577.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1632; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
```

```
SQ   Sequence   341 BP;    73 A;    86 C;    102 G;    68 T;

  Query Match           0.9%;  Score 45;  DB 21;  Length 341;
  Best Local Similarity  75.9%;  Pred. No. 8.92e-10;
  Matches    63;  Conservative    0;  Mismatches  20;  Indels   0;  Gaps   0;

Db    103 agcaagcgncggaaatacgtcttccagtgcacggntggctcctcgtttgtagaagaggtg 162
          |||||||| |||||||||   ||| | ||||| |  |||||||| ||||| || |||||
Qy   4477 agcaagcggcggaaatactctttcgaatgcactgacggctcctcctttgtggacgaggtt 4536

Db    163 gagagacacttagagtgcggctg 185
          |||| |    |  ||||||||||
Qy   4537 gagaaagtggtgaagtgcggctg 4559


RESULT   12
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           0.8%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity  4.6%;  Pred. No. 1.41e-05;
  Matches    5;  Conservative   33;  Mismatches  71;  Indels   0;  Gaps   0;

Db      6 bnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnn 65
          : : : : : : : : : :||  : : : : : : :       :
Cp    153 cctgggcacgctgcgcagccgccagcccgtgacagtccactgtgctgcccgagcaagagca 94

Db     66 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          : : : : : :|||  : : : : : : : : : : : :
Cp     93 ctgcgccgggcacgcctgcggtgccaccttgttcaggatcgccagcact 45


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match           0.8%;  Score 36;  DB 12;  Length 114;
  Best Local Similarity  4.7%;  Pred. No. 4.52e-05;
  Matches    5;  Conservative   32;  Mismatches  69;  Indels   0;  Gaps   0;

Db      6 bnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnn 65
          : : : : : : : : : :||  : : : : : : :       :
Cp    153 cctgggcacgctgcgcagcgccagcccgtgacagtccactgtgctgcccgagcaagagca 94

Db     66 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
          :|||  : : : : : : : : : : : : : : :
Cp     93 ctgcgccgggcacgcctgcggtgccaccttgttcaggatcgccagc 48
```

```
RESULT  14
ID  V44650 standard; DNA; 91 BP.
AC  V44650;
DT  06-OCT-1998  (first entry)
DE  Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW  DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW  anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW  gene therapy; ss.
OS  Mammalia.
PN  WO9827200-A2.
PD  25-JUN-1998.
PF  12-DEC-1997; CA0972.
PR  21-MAY-1997; US-047322.
PR  16-DEC-1996; US-033374.
PA  (UYMC-) UNIV McGILL.
PI  Cossons NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
DR  WPI; 98-362770/31.
PT  Human or mammalian origin of replication consensus sequences - for
PT  inhibiting DNA replication, for controlling initiation of
PT  replication, maintaining circular plasmids and in assembly of human
PT  artificial chromosomes
PS  Claim 1; Page 42; 54pp; English.
CC  This sequence represents a human or mammalian DNA replication origin
CC  consensus sequences of the invention, designated uniorsconsensus.
CC  Administration of the consensus sequence or an anti-gene (comprising a
CC  double stranded copy of the consensus) is used to inhibit DNA replication
CC  in vivo or in vitro. The consensus sequences can also be inserted into an
CC  expression vector, used subsequently for in vitro transfection of
CC  mammalian cells, to control initiation of DNA replication. They can also
CC  be used used to maintain circular plasmids that are capable of
CC  semi-conservative replication in proliferating mammalian cells, or
CC  inserted into mammalian or human artificial chromosome vectors for gene
CC  therapy. Particularly, they are used to create shuttle vector constructs
CC  for defining the essential mammalian elements required for maintenance of
CC  chromosomal function. The consensus sequence can be combined with cloned
CC  human telomeres and large centromeric blocks for assembly of human
CC  artificial chromosomes and maintained as bacterial plasmids, circular or
CC  linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC  elements.
SQ  Sequence  91 BP;  15 A;  1 C;  4 G;  7 T;

  Query Match           0.7%;  Score 32;  DB 46;  Length 91;
  Best Local Similarity  13.8%;  Pred. No. 4.23e-03;
  Matches   11;  Conservative   45;  Mismatches  24;  Indels   0;  Gaps   0;

Db      5 waakrawrwwkkdavwwgakrwwkwvwhrassacmdwkaaktwkggwtwarrywkgrkmw 64
          :||:: ::::  :|:::  : :: : :::|: | ::::| :|::  : :  ::::|: ::
Qy   4679 aaatacagaacagacttatttttattatgagaataaagacttttttttctgcatttggaaa 4738

Db     65 wtwkawsdatakwwwkdakw 84
          : : |: :| | ::: : ::
Qy   4739 aaaaaaaaaaaaaaactcga 4758


RESULT  15
ID  V44650 standard; DNA; 91 BP.
AC  V44650;
DT  06-OCT-1998  (first entry)
DE  Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW  DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW  anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW  gene therapy; ss.
OS  Mammalia.
PN  WO9827200-A2.
PD  25-JUN-1998.
PF  12-DEC-1997; CA0972.
PR  21-MAY-1997; US-047322.
PR  16-DEC-1996; US-033374.
PA  (UYMC-) UNIV McGILL.
PI  Cossons NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
DR  WPI; 98-362770/31.
PT  Human or mammalian origin of replication consensus sequences - for
PT  inhibiting DNA replication, for controlling initiation of
PT  replication, maintaining circular plasmids and in assembly of human
PT  artificial chromosomes
PS  Claim 1; Page 42; 54pp; English.
CC  This sequence represents a human or mammalian DNA replication origin
CC  consensus sequences of the invention, designated uniorsconsensus.
CC  Administration of the consensus sequence or an anti-gene (comprising a
CC  double stranded copy of the consensus) is used to inhibit DNA replication
CC  in vivo or in vitro. The consensus sequences can also be inserted into an
CC  expression vector, used subsequently for in vitro transfection of
CC  mammalian cells, to control initiation of DNA replication. They can also
CC  be used used to maintain circular plasmids that are capable of
CC  semi-conservative replication in proliferating mammalian cells, or
CC  inserted into mammalian or human artificial chromosome vectors for gene
CC  therapy. Particularly, they are used to create shuttle vector constructs
CC  for defining the essential mammalian elements required for maintenance of
CC  chromosomal function. The consensus sequence can be combined with cloned
CC  human telomeres and large centromeric blocks for assembly of human
CC  artificial chromosomes and maintained as bacterial plasmids, circular or
CC  linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC  elements.
SQ  Sequence  91 BP;  15 A;  1 C;  4 G;  7 T;

  Query Match           0.7%;  Score 31;  DB 46;  Length 91;
  Best Local Similarity  9.1%;  Pred. No. 1.27e-02;
  Matches    8;  Conservative   54;  Mismatches  25;  Indels   1;  Gaps   1;

Db      1 awmtwaakrawrwwkkdavwwgakrwwkwvwhrassacmdwkaaktwkggwtwarrywkg 60
          |::|:  :: ::::::: :::| ::: :::::   :| :::: |: :    :|:::::
Cp   1224 atatagggagagaaggttcaagttgtggagatcctga-aaagcatctacccgaaggcagt 1166

Db     61 rkmwwtwkawsdatakwwwkdakwkmwr 88
           ::: |:: :::   : ::::|::: ::
Cp   1165 ttatcttgttggcattcaataataggag 1138


Search completed: Sat May 29 23:08:07 1999
Job time : 1044 secs.
```